

***Image quality evaluation
in X-ray medical imaging
based on Thiel embalmed
human cadavers***

An De Crop

Promotor: Prof. Dr. H. Thierens Co-promotor: Prof. Dr. K. D'Herde
Thesis submitted in fulfillment of the requirements for the degree of
Doctor in Medical Sciences

Department of Basic Medical Sciences
Faculty of Medicine and Health Sciences
2015





Universiteit Gent
Faculteit Geneeskunde en Gezondheidswetenschappen
Vakgroep Medische Basiswetenschappen
Dienst Medische Fysica

Supervisors:

Prof. Dr. H. Thierens (promotor)

Prof. Dr. K. D'Herde (co-promotor)

Guidance Committee:

Prof. Dr. ir. K. Bacher

Prof. Dr. T. Van Hoof

Examination Board

Chairman: Prof. Dr. ir. C. De Wagter¹

Prof. Dr. L. Van Hoorebeke¹

Prof. Dr. G. Villeirs¹

Prof. Dr. ir. B. Vanderstraeten¹

Prof. Dr. P. Parizel²

Dr. M Thijssen³

Prof. Dr. J. Casselman⁴

¹ Ghent University, Ghent, Belgium

² University Hospital Antwerp, Belgium

³ Ziekenhuis Rijnstate, The Nederland

⁴ AZ Sint Jan Brugge-Oostende AV, Belgium

Research institute:

Ghent University

Faculty of Medicine and Health Sciences

Department of Basic Medical Sciences

Medical Physics Division

De Pintelaan 185 B3

B-9000 Ghent, Belgium

Tel.: +32-9-264 65 19

Dankwoord

Het dankwoord, eindelijk... Ik heb er zo naar uitgekeken, en nu weet ik niet waar te beginnen. Ik denk dat ik ga beginnen met: YES!!!!!!!!!!!!!!!!!!!!!! Het waren bijzonder aangename, leerrijke en uitdagende jaren, maar zoals zovelen die ik ken, doe ik dat laatste jaar toch liever niet nog een keer. Maar kom, nu mag ik beginnen aan het meest gelezen deel van een doctoraat: het dankwoord. En in plaats van dat verlossende gevoel, word ik er eigenlijk een beetje melancholisch van, want ondanks de lastige eindsprint, waren het toch 7 fantastische jaren.

Eerst en vooral gaat mijn dank natuurlijk uit naar mijn promotor, Prof. dr. H. Thierens. Tijdens mijn eerste kandidatuur, toen ik nog mocht genieten van “waarde studenten” en “finaal besluiten wij”, had ik nooit gedacht dat u ooit nog de promotor ging worden van mijn doctoraat. Na al die jaren sta ik nog steeds versteld van al uw kennis. Ik ben dan ook ontzettend dankbaar dat ik de mogelijkheid heb gekregen om gebruik te maken van dit vat vol informatie. Zeker dit laatste jaar (toen u al iets minder tijd moest spenderen aan de “waarde studenten”), heb ik bijzonder veel bijgeleerd uit de constructieve discussies en uw avondlijke/nachtelijke thesisverbeteruurtjes hebben ervoor gezorgd dat ik de deadline gehaald heb. Ik hoop dat u tijdens uw emeritaat dan ook de tijd vindt voor de dingen waar u zo van geniet en wie weet komen we elkaar nog wel eens tegen op een (niet al te hoge) berg.

Het is het laatste jaar veel ter sprake gekomen: eigenlijk ben je ‘niets’ van mij, maar ik vrees dat er zonder jou ook niets van dit doctoraat in huis zou zijn gekomen. Dus daarom Klaus, of in het lang, Prof. Dr. ir. K. Bacher, bedankt!! In het begin van mijn doctoraat heb je mij gezegd: “wie het hardste zaagt, die wordt eerst geholpen”. Nu hoor ik mij te excuseren voor het vele zagen, maar dat ga ik niet doen. Mocht ik opnieuw moeten beginnen, ik zou waarschijnlijk nog veel harder zagen. Het is namelijk ongelooflijk hoeveel je van ‘straling’ weet. Ik ben dan ook bijzonder blij dat ik het geluk heb gehad om, ondanks onze titelloze relatie, door u begeleid te

worden. Ik heb ontzettend veel van u geleerd, maar wees gerust, het is niet omdat ik vertrek uit Gent, dat ik niet meer kan/zal zagen.

Ook mijn copromotor, Prof. Dr. K. D'Herde, verdient een woord van dank. De samenwerking met de afdeling anatomie was onontbeerlijk voor dit doctoraat. In dit kader wil ik zeker ook Prof. Dr. T. Van Hoof bedanken. Tom, ettelijke uren hebben we samen al scannend doorgebracht, soms tot in de nachtelijke uurtjes. We hebben massa's roze kaderkes gezien, versleept en er ook soms op gevloekt. Gelukkig krijgen we die ondertussen wel al op de juiste plaats. Bedankt voor de enorm vlotte samenwerking, voor je goede humeur, de babbels en voor je perfecte stem-imitaties van de plaatselijke hoogwaardigheidsbekleders. De (ex)-medewerkers van de dienst anatomie verdienen zeker ook een woord van dank: Aron, Michael, Isabelle: bedankt voor al het gezeul en voor de vlotte samenwerking.

Bedankt aan de leden van de examencommissie. Jullie feedback en kritische opmerkingen hebben mij geholpen om het onderste uit de kan te halen.

Als je een doctoraat wil schrijven over de link tussen klinische en fysisch-technische beeldkwaliteit, dan is de samenwerking met radiologen natuurlijk cruciaal. Vooral de input van Dr. Smeets en Dr. Casselman hebben mij enorm vooruit geholpen. Dr. Smeets, u heeft onnoemelijk veel beelden voor mij uitgelezen. Met veel geduld en enthousiasme heeft u mij geholpen met het zoeken naar voorbeelden op de PACS, het aanduiden van Chinese vrijwilligers, het zoeken naar de juiste kwaliteitscriteria. En dankzij u herken ik fissuren, micronodulaire structuren en laten we vooral de pneumothorax niet vergeten. Dr. Casselman, ik had nooit gedacht dat ik de metaalartefactenstudie nog in mijn doctoraat zou kunnen verwerken, maar dankzij uw medewerking, enthousiasme, gedrevenheid en enorme kennis verliep deze studie als een trein.

Ook alle assistenten die beelden hebben uitgelezen mogen zeker niet ontbreken in dit dankwoord: Merel Vergauwen, Barbara Smet, Ulla Kiendys, Mathias Van Borsel, Tom Dewaele, Nicolas Bossu, Elke Vereecke en Jaime Pamplona.

De laatste studie van dit doctoraat zou nooit zo vlot verlopen zijn zonder de medewerking van Melissa Dhierens. Melissa, voor QCC werkten we altijd al vlot samen, maar het was zeer leuk om eens te mogen samenwerken voor het onderzoek ook. Veel succes met de verdere loopbaan en natuurlijk ook met de kindjes.

Als doctoraatsstudent krijg je natuurlijk thesisstudenten toegewezen. Het was altijd een beetje afwachten of er niemand ging flauwvallen, maar jullie hebben de eerste kennismaking met Thiel lichamen allemaal met glans doorstaan. Siska (ook al was je officieel dan nog niet mijn thesisstudent), Jolien, Eline en Monica: bedankt voor het sorteren van beeldjes, het tekenen van ROI's en de talloze analyses in ellenlange excelfiles.

Als er een groep is die dit dankwoord zeer aandachtig zal lezen, dan zijn het wel de collega's. Mijn eerste INW ervaring dateert al van 2005, toen ik voor mijn thesis van de biomedische wetenschappen in de wandelgangen rondliep. Na nog een thesis, een stage en een vervangingsopdracht, ging ik in oktober 2008 dan echt bij jullie aan de slag en kreeg ik voor de duidelijkheid toch nog eens uitgelegd dat er 2 verschillende diensten op het INW zitten: de fyco en de medische fysica. Voor de veiligheid zal ik hier dus ook maar het onderscheid maken. Aan de (ex)-collega's van de fyco: Isabelle, Myriam, Nancy, Levi, Lothar, Robbe en Liesbeth, bedankt voor de vele babbels tijdens de middagpauzes. Nele, bedankt voor de talloze onnozele computerproblemen die je voor mij hebt opgelost, maar vooral ook bedankt voor de opbeurende gesprekken.

De collega's van de medische fysica – Kim, Virginie en Régine – bedankt voor de praktische ondersteuning en de fijne babbels. Régine, voor het doctoraat hebben we niet veel samengewerkt, maar voor het onderwijs des te meer. Bedankt, je hebt mij ontzettend veel geleerd. Ik hoop dat je van de welverdiende rust kan genieten samen met je man, kinderen en de nog steeds groter wordende groep kleinkinderen. Oud-collega's Joke en Iris maar zeker ook Laurence en Sofie, het was fijn samenwerken, vooral de handige tips en tricks over de eindfase van het doctoraat zijn goed van pas gekomen. Charlot, officieel was je niet mijn thesisstudent, maar officieus zal ik je toch beschouwen als 1 van mijn eerste studenten. Jouw goede humeur, je gedrevenheid maar vooral je enthousiasme voor het onderzoek zullen mij zeker bijblijven. Ik wens je ontzettend veel succes in Zuid-Afrika, maar ik ben ervan overtuigd dat je ze daar, net zoals je hier gedaan hebt, omver zal blazen. QCC genoten Caro en Lore, amai wat hebben wij veel gelachen!!! Ik kan mij eerlijk gezegd echt niet inbeelden dat ik nog zulke hilarische conversaties ga voeren met mijn nieuwe collega's, maar ik hou jullie in elk geval op de hoogte. Bedankt voor het uitlezen van de bolletjes, voor de versterking in het QCC-team toen het echt nodig was en voor de feedback op alle al dan niet doctoraatgerelateerde perikelen. Dimitri, je bent er nog maar net bijgekomen, maar je hebt je al bijzonder goed ingewerkt, je staat met gemak je mannetje tussen het vrouwelijk QCC bastion en je "burgie-kennis" is mij al zeer goed

van pas gekomen. Ik wens jullie alle drie bijzonder veel succes voor het vervolg van jullie doctoraat en hoop stiekem toch dat ik nog word uitgenodigd op de verdediging. En Lore, als mijn assistent-opvolgster wens ik je natuurlijk veel succes met de studenten. Laat je niet kisten door de LO'ers, maar als je jouw triatlon resultaten vermeldt, zullen ze wel rap zwijgen denk ik.

De bureaugenoten, An en Liesbeth. An, tijdens mijn stage kon je mij omver blazen met HVL's, AEC's en DICOM headers. Maar dankzij jouw vakkundige begeleiding, kan ik ondertussen al naarstig meedoen in het omverblazen. Het was je nooit teveel moeite om mee te helpen zoeken naar een oplossing voor de talloze kleine en grote problemen die zijn opgedoken tijdens mijn doctoraat. Zonder aarzelen schoof je je eigen grandioze stapel werk aan de kant om op zoek te gaan naar de perfecte oplossing/uitleg. Hoe jij erin slaagt om alles te combineren en om daarbovenop prachtige fotoalbums en knutselwerkjes te fabriceren, het zal mij altijd een raadsel blijven.

Liesbeth, zelfs toen we nog samen studeerden in 2007 en 2008 vonden de mensen het bizar dat we elkaar niet beschouwden als concurrenten. Toen je dan in 2010 bij ons kwam werken vroegen ze: "ja maar ja, is dat wel een goed idee, samenwerken met een vriendin". En toen we dit jaar alle twee op zoek moesten naar een job was het weer van: "oei oei, in concurrentie met elkaar, zou er geen breuk komen in de vriendschap". Of het nu in 2007 was of in 2015, ik heb mij eerlijk gezegd nooit zorgen gemaakt. Ik ben ervan overtuigd dat wij oprecht het beste met elkaar voorhebben en elkaar dus ook het beste gunnen. Ik kan niet beschrijven hoe blij ik ben dat ik dit volledige doctoraatsavontuur met jou heb kunnen delen. Maar meer nog vind ik het fantastisch dat we al de rest ook hebben kunnen delen, van mening geven over keukeninrichting tot het blij in de lucht springen bij het melden van (nog) een zwangerschap. Het doctoraatsverhaal is dan voor ons alle twee eindelijk afgelopen, maar wij functioneren nu eenmaal beter in duo, dus je bent nog niet van mij vanaf ☺.

De vrienden, al dan niet stralend ☺, bedankt voor de lunches, brunches en dinerdates. De gemeenschappelijke klaaguurtjes hebben deugd gedaan!

De familie en schoonfamilie: Katrijn en Hans, Stijn en Liesebet, Maaïke en Bram, Opa en Oma, Mama en Raf, Michiel en Tine, Matthijs en Aijia, Eva en Elie, wat kan ik zeggen, ik ben fan van grote families!! Bedankt voor alles, het babysitten, het luisterende oor, de interesse, de spellingscontrole, de vele telefoontjes...

Vake en moeke, ik weet eerlijk gezegd niet wie het meeste stress heeft gehad voor mijn doctoraat, jullie of ik. Toen ik in 2005 schoorvoetend kwam vragen of ik aub mocht gaan samenwonen met Thomas, was jullie enige bezorgdheid mijn studies. Tien jaar later kan ik jullie met bijzonder veel trots zeggen dat er niets is om jullie zorgen over te maken, alhoewel ik sterk betwijfel of jullie er ooit in gaan slagen om jullie geen zorgen te maken over de kinderen. Van jullie erfde ik het o zo belangrijke doorzettingsvermogen en het lef om er te geraken, bedankt voor alles!!

Thomas, velen beschouwen een doctoraat als een eenzame rit, maar ik ben het daar niet mee eens. Waar je ook was, je was altijd maar een telefoontje van mij verwijderd en onze gsm factuur zal dat geweten hebben. Je nuchtere kijk op de zaak, je zalige humor en je geloof in mij maakten dit allemaal zoveel gemakkelijker. Het is ondertussen al meer dan 13 jaar geleden dat ik je heb gevraagd: "ben je van plan met mijn voeten te rammelen, of hoe zit dat?". Vier huizen, twee thesissen, 1 doctoraat, 1 verbouwing en twee kinderen later, is het antwoord op die vraag wel duidelijk: JA! Ik denk dat je er ondertussen uw levensdoel van gemaakt hebt om mij op mijn paard te krijgen en ik geniet nog steeds van de twinkellichtjes in je ogen als je weer eens in je opzet geslaagd bent. We hebben al veel uitdagingen samen overwonnen dus ik denk dat we die nieuwe job en spookje nummer drie ook zonder problemen zullen overleven.

Hanne en Lucas, mijn allerliefste engeltjes maar zeker ook mijn allergrootste spookjes. Niets beter dan kinderlijke onschuld om mij altijd terug met mijn twee voeten op de grond te zetten en de relativiteit van alles te doen inzien. Jullie onvoorwaardelijke liefde is goud waard.

BEDANKT!!!

An

September 2015

Contents

Summary	i
Samenvatting	v
List of acronyms	ix
PART I : INTRODUCTION	1
1 THE NEED FOR OPTIMIZATION IN DIAGNOSTIC RADIOLOGY	3
2 IMAGE QUALITY ASSESSMENT	7
2.1 Basic quality parameters in X-ray imaging	7
2.2 Physical-technical image quality analysis	8
2.3 Patient models	12
2.3.1 Anthropomorphic phantoms	12
2.3.2 Thiel embalmed cadaver	15
2.4 Clinical image quality analysis	17
2.4.1 Receiver Operator Characteristics (ROC)	17
2.4.2 Visual Grading Analysis (VGA)	20
2.5 Observer Models	20
references	24
PART II: AIM AND OUTLINE OF THE THESIS	29
3 AIM OF THE THESIS	31
4 OUTLINE OF THE THESIS	33
references	35
PART III: ORIGINAL RESEARCH	37
5 PAPER I	39
5.1 Background	39
5.1.1 Digital radiography	39
5.1.2 Thorax radiography	41

5.2	Paper I	42
6	PAPER II	57
6.1	Background	57
6.1.1	CT	57
6.1.2	Iterative reconstruction	58
6.1.3	Thorax CT	61
6.2	Paper II	63
7	PAPER III	83
7.1	Background	83
7.1.1	Dual energy CT	83
7.1.2	Metal artifacts	85
7.1.3	Head and neck CT	86
7.2	Paper III	87
	references	105
	PART IV: GENERAL DISCUSSION	107
8	STRENGTH AND WEAKNESS OF DIFFERENT IMAGE QUALITY PARAMETERS	113
8.1	Physical-technical image quality parameters	113
8.1.1	Noise	113
8.1.2	Resolution	115
8.1.3	Contrast	116
8.1.4	Contrast-detail	116
8.2	ROC and VGAS	118
8.3	Model observers	120
9	THE VALUE OF THIEL EMBALMED HUMAN CADAVERS	123
10	CLINICAL VERSUS PHYSICAL IMAGE QUALITY ASSESSMENT	127
11	FINAL CONCLUSIONS	133
12	FUTURE PERSPECTIVES	137
	references	133
	PART V : CURRICULUM VITAE – AN DE CROP	137

Summary

Image quality analysis represents a crucial component in the radiation dose optimization process in medical imaging. Physical-technical image quality parameters focus more on the performance of the imaging equipment and do not address the clinical performance of the image modality. Clinical image quality describes the effectiveness at which an image can be used for its intended purpose. However, the link between clinical and physical-technical image quality is still an open question. The aim of this PhD dissertation is to assess the image quality of different clinical X-ray applications using both clinical and physical image quality parameters. Moreover, the correlation between physical-technical and clinical image quality parameters is investigated.

Chest radiography is one of the most frequently performed diagnostic X-ray examinations. The frequent use and diagnostic importance of chest X-ray makes it an interesting topic for optimization of image quality and patient dose. Therefore, the correlation between clinical and physical-technical image quality was assessed for digital chest radiography. Clinical image quality of chest radiographs was assessed by means of three human cadavers, conserved using the Thiel embalming technique. Physical-technical image quality was assessed using the CDRAD 2.0 contrast-detail phantom. Four experienced radiologists assessed the image quality, using a visual grading analysis (VGA) technique based on the European Quality Criteria for Chest Radiology. The CDRAD images were scored manually and automatically using dedicated software, both resulting in an Image Quality Figure inverse (IQF_{inv}) value. For both manually and automated obtained IQF_{inv} values, a statistically significant correlation ($r = 0.80$, $P < 0.01$ and $r = 0.92$, $P < 0.001$, respectively) with the VGA scores was observed, supporting the value of contrast-detail phantom analysis for evaluating clinical image quality in chest radiography.

As the first study showed an excellent correlation between clinical and physical-technical image quality in digital chest radiography, the correlation in chest CT was investigated in the

second study. To this end, the potential dose reduction of iterative reconstruction (IR) compared to conventional filtered back projection based on different clinical and physical-technical image quality parameters was assessed. Clinical image quality was assessed using three Thiel embalmed human cadavers. Four experienced radiologists assessed the image quality, using a VGA technique based on the European Quality Criteria for Chest Tomography. A Catphan phantom was used to assess physical-technical image quality parameters such as noise, contrast-detail and contrast-to-noise ratio (CNR). Correlation coefficients between clinical and physical-technical image quality varied from 0.88 to 0.92, depending on the selected physical-technical parameter. Depending on the strength of SAFIRE, the potential dose reduction based on noise and CNR and IQF_{inv} varied from 14.0% to 67.8%, 16.0% to 71.5% and 22.7% to 50.6% respectively. Potential dose reduction based on clinical image quality varied from 27% to 37.4%, depending on the strength of SAFIRE. Our results demonstrate that noise assessments in a uniform phantom overestimate the potential dose reduction for the SAFIRE IR algorithm.

A variety of tools is available to diminish dental metallic artifacts in head and neck CT imaging. However, up till now, no comparison of the effect of different tools was made. Therefore, the diminishing effect on artifacts of the following techniques was studied: altering kVp, application of model based iterative reconstruction (Veo), use of virtual monochromatic images and metal artifact reduction software (MARs). Clinical image quality was assessed using a Thiel embalmed cadaver. Four experienced radiologists assessed the image quality, using a VGA technique. Physical-technical image quality was assessed in a uniform polymethylmethacrylate (PMMA) cylindrical phantom and in the Catphan phantom. Significant influence of increasing kVp and the use of Veo was found on clinical image quality ($p = 0.007$ and $p = 0.014$, respectively). Application of MARs resulted in a smaller artifact area ($p < 0.05$). However, MARs reconstructed images resulted in lower clinical image quality. Of all investigated techniques, Veo shows to be most promising, with a significant improvement of both the clinical and physical-technical image quality without adversely affecting contrast detail. MARs reconstruction in CT images of the oral cavity to reduce dental hardware metallic artifacts appeared not to be sufficient, and may even adversely influence the image quality.

Both physical-technical and clinical image quality parameters are very important tools to assess the image quality. However, it is important to understand the strengths and weaknesses of both methods. Physical technical measurements are fairly easy to conduct in

quality assurance programs. They are essential for evaluating equipment performance and acceptance and constancy testing. However, for information about the actual clinical performance related to pathological anatomy like lesion detection, clinical assessment studies based on ROC studies still need to be performed. Clinical image quality assessment is rather difficult to perform in routine practice. However, they are the only reliable option for optimization purposes.

For some specific conditions like chest radiography, an excellent correlation was found between the results of a contrast-detail study and the assessment of the normal anatomy. This was not the case for chest CT imaging. The data of the PhD study indicate that efforts should be made for construction of more sophisticated contrast-detail phantoms for CT imaging to improve the correlation for these imaging modalities.

Samenvatting

De analyse van beeldkwaliteit is een cruciale component bij het optimaliseren van de stralingsdosis in de medische beeldvorming. Fysisch-technische beeldkwaliteitsparameters focussen zich meer op de eigenschappen van het toestel maar houden geen rekening met de klinische kwaliteit van de beeldvormingsmodaliteit. Klinische beeldkwaliteit beschrijft de efficiëntie waarmee een beeld voor zijn doel kan gebruikt worden. De link tussen klinische en fysische beeldkwaliteit is echter nog steeds een onbeantwoorde vraag. Het doel van deze PhD-thesis is het beoordelen van de beeldkwaliteit voor verschillende klinische X-stralentoepassingen, gebruikmakende van zowel klinische als fysisch-technische beeldkwaliteitsparameters. Bovendien wordt ook de correlatie tussen fysisch-technische en klinische beeldkwaliteitsparameters onderzocht.

Thoraxradiografie is één van de meest frequent uitgevoerde diagnostische X-stralenonderzoeken. Het frequent gebruik en de diagnostische waarde van thoraxradiografie maken het een interessant onderwerp voor optimalisatie van zowel beeldkwaliteit als patiëntdosis. Daarom werd de correlatie tussen klinische en fysisch-technische beeldkwaliteit geanalyseerd bij thoraxradiografie. De klinische beeldkwaliteit van thoraxradiografieën werd beoordeeld met behulp van drie humane kadavers die gebalsemd werden volgens de methode van Thiel. Het CDRAD 2.0-fantoom werd gebruikt voor de beoordeling van de fysisch-technische beeldkwaliteit. Vier ervaren radiologen beoordeelden de klinische beeldkwaliteit op basis van een Visual Grading Analysis (VGA) gebaseerd op de Europese kwaliteitscriteria voor thoraxradiografie. Het CDRAD-fantoom werd zowel manueel als automatisch uitgelezen waardoor in beide gevallen een inverse Image Quality Figure-waarde (IQF_{inv}) wordt bekomen. Voor zowel de manuele als de automatische bekomen IQF_{inv} -waardes wordt een significante correlatie gevonden met de VGA-score (respectievelijk $r = 0.80$, $P < 0.01$ en $r = 0.92$, $P < 0.001$).

Dit ondersteunt de waarde van contrastdetailanalyse voor de evaluatie van klinische beeldkwaliteit in thoraxradiografie.

Aangezien in de eerste studie een uitstekende correlatie werd aangetoond tussen klinische en fysisch-technische beeldkwaliteit in digitale thoraxradiografie, werd deze correlatie in een tweede studie onderzocht voor thorax CT-onderzoeken. Hiervoor werd de potentiële dosisreductie bij gebruik van iteratieve reconstructie (IR) in vergelijking met filtered back projection onderzocht aan de hand van verschillende klinische en fysisch-technische beeldkwaliteitsparameters. Drie Thiel-gebalsemde humane kadavers werden gebruikt voor de beoordeling van de klinische beeldkwaliteit. Vier ervaren radiologen beoordeelden de beeldkwaliteit gebruik makende van een VGA-techniek gebaseerd op de Europese kwaliteitscriteria voor thorax-CT. Het Catphan fantoom werd gebruikt voor de beoordeling van fysisch-technische beeldkwaliteitsparameters zoals ruis, contrastdetail en contrast-to-noise ratio (CNR). De correlatiecoëfficiënten tussen de klinische en fysisch-technische beeldkwaliteit varieerden van 0.88 tot 0.92, afhankelijk van de geselecteerde fysisch-technische parameter. Afhankelijk van de sterkte van SAFIRE varieerde de potentiële dosisreductie gebaseerd op ruis, CNR en IQF_{inv} respectievelijk van 14.0% tot 67.8%, 16.0% tot 71.5% en 22.7% tot 50.6%. De potentiële dosisreductie gebaseerd op de klinische beeldkwaliteit varieerde van 27% tot 37.4% afhankelijk van de sterkte van SAFIRE. Onze resultaten tonen aan dat ruisbeoordeling in een uniform fantoom de potentiële dosisreductie voor het SAFIRE-IR-algoritme overschat.

Voor het reduceren van dentale metaalartefacten in hoofd en nek CT-beeldvorming bestaan er verschillende technieken. Tot nu toe is er echter nog geen vergelijking van het effect van de verschillende technieken gemaakt. Daarom werd de reducerende invloed op metaalartefacten van aanpassing van de kVp, de toepassing van modelgebaseerde iteratieve reconstructie (Veo), het gebruik van virtueel monochromatische beelden en metaal artefact reductie software (MARs) onderzocht. Een Thiel-gebalsemd humaan kadaver werd gebruikt voor de analyse van de klinische beeldkwaliteit. Hiervoor beoordeelden vier ervaren radiologen de beeldkwaliteit aan de hand van een VGA techniek. Fysisch-technische beeldkwaliteit werd beoordeeld in een uniform cilindrisch polymethylmethacrylaat-fantoom (PMMA) en een Catphan-fantoom. Stijging in kVp en het gebruik van Veo resulteerden in een significante invloed op de klinische beeldkwaliteit (respectievelijk $p = 0.007$ en $p = 0.014$). Toepassing van MARs resulteerde in een kleinere artefactenregio ($p < 0.05$). Er werd echter een lagere klinische beeldkwaliteit bekomen bij MARs gereconstrueerde beelden. Van alle onderzochte technieken blijkt Veo de

meest belovende met een significante verbetering van zowel de klinische als de fysisch-technische beeldkwaliteit zonder dat contrastdetail nadelig wordt beïnvloed. Het gebruik van MARs gereconstrueerde CT-beelden van de mondholte blijkt de dentale metaalartefacten onvoldoende te reduceren en kan leiden tot een negatieve beïnvloeding van de beeldkwaliteit.

Zowel fysisch-technische als klinische beeldkwaliteitsparameters zijn zeer belangrijke tools in de beoordeling van beeldkwaliteit. Het is echter belangrijk om de sterktes en zwaktes van de beide methodes te verstaan. Fysisch-technische metingen zijn redelijk gemakkelijk om uit te voeren in kwaliteitborgingsprogramma's. Ze zijn essentieel voor de evaluatie van toesteleigenschappen en acceptatie- en stabiliteitstesten. Voor informatie omtrent de klinische prestatie gerelateerd met de pathologische anatomie zoals lesiedetectie moeten nog klinische beoordelingen zoals ROC-studies uitgevoerd worden.

Voor enkele specifieke onderzoeken zoals thoraxradiografie werd een uitstekende correlatie gevonden tussen de resultaten van een contrastdetailstudie en de beoordeling van de normale anatomie. Deze correlatie werd niet gevonden voor thorax CT-beeldvorming. De data van deze PhD-thesis tonen aan dat er nood is aan de constructie van een meer geavanceerd contrastdetailfantoom voor CT-beeldvorming om de correlatie voor deze modaliteiten te verbeteren.

List of acronyms

A

AFROC	Alternative Receiver Operator Characteristics
AI	Artifact Index
ASIR	Adaptive Statistical Iterative Reconstruction
AUC	Area Under the Curve

C

CBCT	Cone Beam Computed Tomography
CD	Contrast-Detail
CEC	Commission of the European Communities
CNR	Contrast to Noise Ratio
CoCr	Cobalt-Chromium
CR	Computed Radiography
CT	Computed Tomography
CTDI	Computed Tomography Dose Index

D

DE	Dual Energy
DECT	Dual Energy Computed Tomography
DR	Direct Radiography
DRL	Dose Reference Level
DQE	Detective Quantum Efficiency

F

FANC	Federaal Agentschap voor Nucleaire Controle
FBP	Filtered Back Projection
FFE	Free response Forced Error
FROC	Free response Receiver Operator Characteristics

I

ICC	Intraclass Correlation Coefficient
ICRP	International Commission on Radiological Protection
ICRU	International Commission on Radiation Units
ICS	Image Criteria Score
IQ	Image Quality
IQF	Image Quality Figure
IR	Iterative Reconstruction
IRIS	Iterative Reconstruction in Image Space

M

MAFC	Multiple Alternative Forced Choice
MARs	Metal Artifact Reduction software
MBIR	Model Based Iterative Reconstruction
MRI	Magnetic Resonance Imaging
MTF	Modulation Transfer Function

N

NPS	Noise Power Spectrum
-----	----------------------

P

PMMA	Polymethylmethacrylate
PSF	Point Spread Function

R

ROC	Receiver Operator Characteristics
ROI	Region Of Interest

S

SAFIRE Sinogram Affirmed Iterative Reconstruction

SNR Signal to Noise Ratio

T

Ti Titanium

TLD Thermoluminescence dosimeter

U

UNSCEAR United Nations Scientific Committee on the Effects of Atomic Radiation

V

VGA Visual Grading Analysis

VGAS Visual Grading Analysis Score

Z

Zr Zirconium

Part I:

Introduction

The need for optimization in diagnostic radiology

Diagnostic radiology generally refers to the analysis of images obtained using X-rays, ultrasound and magnetic resonance imaging (MRI). X-ray diagnostic radiology includes plain radiography, dental radiology, fluoroscopy, interventional radiology, computed tomography (CT) and Cone Beam Computed Tomography (CBCT) [1].

With improved health care, increasing availability of medical equipment and the aging of the population, the number of radiological medical procedures is increasing considerably. According to the UNSCEAR 2008 Report [1] the annual frequency of diagnostic medical examinations (excluding dental radiology) in developed countries (health-care level I) is estimated to have increased from 820 per 1 000 population in 1970-1979 to 1 332 per 1 000 population in 1997-2007 (*Figure 1-1*). Worldwide, the use of diagnostic medical radiological examinations increased from 330 per 1 000 populations in 1991-1996 to 488 per 1 000 in 1997-2007. Correspondingly, worldwide, the estimated annual per caput dose from diagnostic medical examinations has increased from 0.40 mSv to 0.62 mSv [1]. More recent data (2011) from the European Dose Datamed 2 project report a mean European per caput effective dose of 0.98 mSv and a mean Belgian per caput effective dose of 1.96 mSv [2] (*Figure 1-2*).

The contribution of CT examinations to the population dose has increased rapidly since this medical imaging modality was introduced in the 1970s. As a consequence, the increasing trend in annual CT examination frequency and the significant dose per examination have an important impact on the overall population dose due to medical exposures.

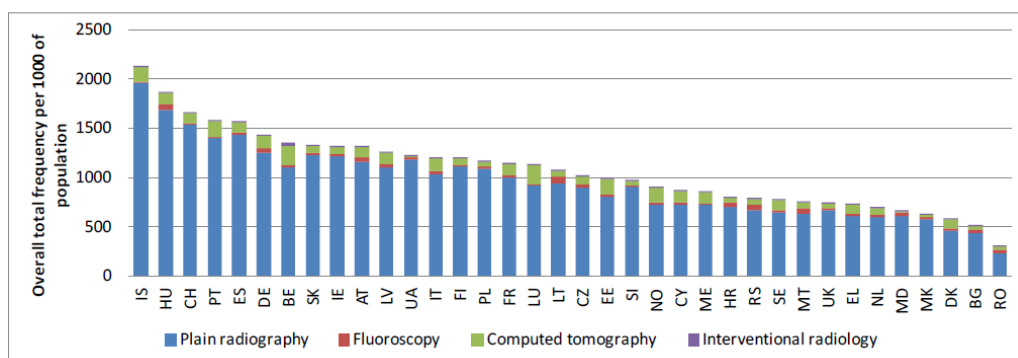


Figure 1-1: Overall total frequencies per 1000 of population for different European countries. The relative contributions of the four main groups (plain radiography including dental, fluoroscopy, computed tomography and interventional radiology) are also shown [2].

In the UNSCEAR 2000 Report [3], it was noted that worldwide, for the period 1991 to 1996, 34% of the collective dose due to medical exposures was related to CT examinations. For the period 1997-2007, worldwide, CT already accounts for 43% of the total effective dose [1]. For the same period, CT scanning accounts for 7.9% of the total number of diagnostic medical examinations in developed countries. However, the contribution of CT scanning to the total collective effective dose due to diagnostic medical examinations in developed countries is approximately 47% [1].

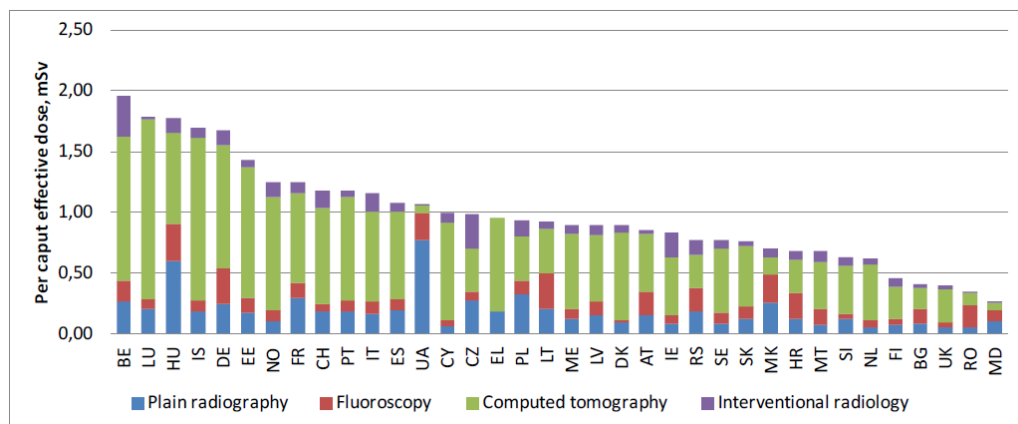


Figure 1-2: Per caput effective doses for different European countries. The relative contributions of the four main groups (plain radiography, fluoroscopy, computed tomography and interventional radiology) are also shown. [2]

According to the European Dose Datamed 2 report, the relative frequency of CT as a percentage of the overall frequency of all X-ray examinations is 8.7% while the relative

contribution to the overall collective effective dose is 55% [2] (*Figure 1-3*). In Belgium, the use of CT increased from 1.4 million in 1994 to 2.1 million in 2013 [4].

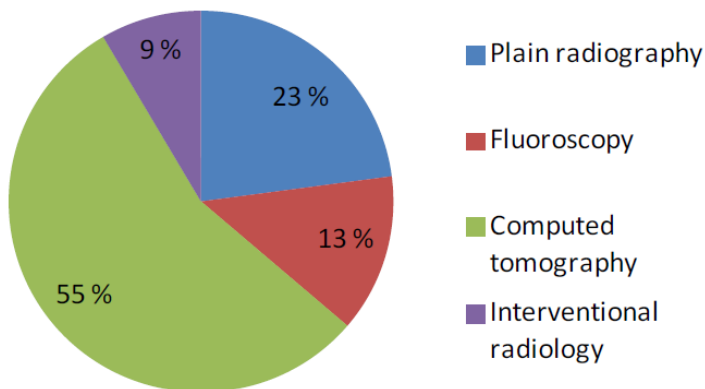


Figure 1-3: Contribution of the four main groups to the overall collective effective dose [2].

The increasing trend in both annual frequency and annual per caput effective dose of diagnostic medical radiological examinations combined with the on-going technological developments of radiological equipment, emphasize the need for dose evaluation and optimization. Both the International Commission on Radiological Protection (ICRP) [5] and the European medical exposure directive [6] recommend the optimization of patient dose in diagnostic radiology.

A part of the optimization process is the establishment of diagnostic reference levels (DRL). DRL's were first mentioned by the ICRP in 1990 [7] and subsequently recommended in greater detail in 1996 [5]. Reference levels are typically set at the 75th percentile of the dose distribution from a survey conducted in different national or international hospitals and are helpful in identifying potentially unusual practice (the highest 25% of typical doses). The 25th percentile of the dose distribution can be regarded as the reference level of good practice. DRL's are a practical tool to promote the assessment of existing protocols and enable the development of new and improved protocols at each center by highlighting the substantial variations in practice between different hospitals for similar types of examinations and similar patient groups. They are intended to promote awareness, dose audit and comparison as the base for improving patient radiation protection. Each center should determine its typical mean dose levels and compare these with the relevant DRL. Mean values above the DRL should be investigated and either justified as being clinically necessary or reduced through appropriate changes in practice to improve patient protection.

Dose surveys should be repeated periodically to establish new reference levels. Data of the Federal Agency of Nuclear Control (FANC) show a decrease over time of the DRL's of most examinations [8]. However, the image quality is not assessed during the process of determining the DRL's. Though, the radiation dose is inextricably related to the image quality. Therefore, unrestricted reduction of the radiation dose is impossible without compromising the image quality. Consequently, image quality analysis represents a crucial component in the radiation dose optimization process.

Image quality assessment

2.1 Basic quality parameters in X-ray imaging

Different physical parameters can be used to describe the image quality. In basic medical physics, image quality is determined by noise, spatial resolution and contrast.

In radiographic imaging, *noise* refers to unwanted image details that interfere with the visualization of an abnormality of interest and with the interpretation of an image. There are usually two major sources of noise: anatomic noise and radiographic noise [9]. The former refers to normal unwanted anatomic variations within an image (e.g., the rib projection pattern in a chest radiograph confounding the detection of a lung nodule). As such, the characterization of anatomic noise is task-dependent and is not directly related to the intrinsic performance of a detector. The second type of noise, radiographic noise, refers to unwanted variations within an image that do not originate within the image subject. Generally speaking, radiographic noise is composed of two types: quantum noise and system noise [10]. Quantum noise is caused by the limited number of photons forming the image. When an average number of X-rays in a pixel is N , the noise σ in the pixel, defined as the standard deviation of N , will be:

$$\sigma = \sqrt{N} \text{ [11]}$$

System noise is added by the additional statistically random and systematic processes in the detector during image formation. Radiographic noise measurements can easily be performed by imaging a uniform phantom and determining the standard deviation of the pixel values.

Technically, the *resolution* of a system is the minimum distance between two objects necessary to still be distinguished as distinct objects in the image. This is frequently described in visually discernible line pairs per mm (lp/mm).

Contrast is the difference in image grayscale between closely adjacent regions on the image. Contrast is generated by the differential attenuation of X-rays in tissues and is affected by the applied X-ray energy spectrum and the contrast resolution capabilities of both detector and display system [11].

2.2 Physical-technical image quality analysis

To describe the above mentioned fundamental physical characteristics, different quantities can be analyzed.

Radiographic noise is best characterized by its Noise Power Spectrum (NPS). The NPS is the variance of noise within an image divided among various spatial frequency components of the image [12]. It indicates how an imaging system passes noise from the input to the output. Because radiographic noise does not include anatomic variations, the appropriate image for a NPS analysis is a uniform flat exposure with no object in the field of view.

Relative noise, represented by the Signal to Noise Ratio (SNR), refers to the magnitude of image fluctuations relative to the signal present in the image. For a given number of detected photons, N , and a radiographic noise, σ , the maximum available SNR is:

$$SNR = \frac{S}{\sigma} = \frac{N}{\sqrt{N}} = \sqrt{N} \quad [12]$$

In an ideal imaging device, the only source of noise is associated with the finite number of X-ray photons forming the image, i.e., quantum noise. Quantum noise is related to the Poisson statistics thus the SNR^2 of an ideal imaging system is proportional to the number of individual X-ray quanta forming the image.

$$SNR_{ideal} = \sqrt{N} \quad [12]$$

However, devices used in radiography are not ideal, they contain sources of noise not related to the Poisson-distributed X-ray flux. Thus, the SNR measured in such a non-ideal device ($SNR_{non-ideal}$) contains noise properties worse than noise due to X-ray quantum statistics alone. Because of detector inefficiencies, the $SNR_{non-ideal}$ is always smaller than the SNR_{ideal} . The ratio of the $SNR_{non-ideal}$ to the SNR_{ideal} can be used to define the efficiency of the system. This ratio is called the detective quantum efficiency (DQE).

$$DQE = \frac{SNR_{non-ideal}^2}{SNR_{ideal}^2} = \frac{SNR_{non-ideal}^2}{N} \quad [12]$$

The DQE describes how efficiently a system translates incident X-ray photons into useful signal within an image.

A thorough description of the system's spatial resolution is given by the point spread function (PSF). The PSF describes the response of an imaging system to a point stimulus or bead in the spatial domain (*Figure 2-1*). The PSF is also an attractive method for characterizing the CT scanner because the PSF from the bead can be used to describe the spreading in the conventional axial (x,y) plane of the CT scanner and the spreading of the point in the slice thickness or Z domain [13]. This combination of information from x, y and z axes, essentially describes the 3D PSF.

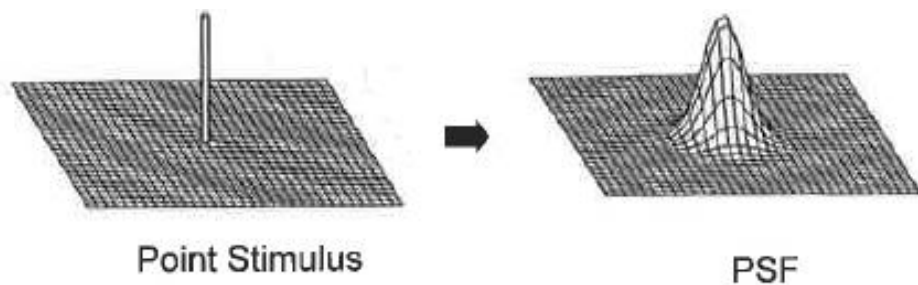


Figure 2-1: Display of a point stimulus input to an imaging system and the response of an imaging system to that point stimulus, PSF [11]

Another useful way to express the resolution of an imaging system is to make use of the frequency domain. The Modulation Transfer Function (MTF) of an image system, is a very complete description of the resolution properties of an imaging system. The MTF is a plot of the imaging system's modulation versus spatial frequency [11]. The MTF illustrates the

fraction (or percentage) of an object's contrast that is recorded by the imaging system, as a function of the size (i.e. spatial frequency) of the object (*Figure 2-2*). Many imaging systems are really imaging chains, where the image passes through many different intermediate steps from the input to the output of the system. To understand the role of each component in the imaging chain, the MTF is measured separately for each component. The total system MTF at any frequency is the product of all the subcomponent MTF curves.

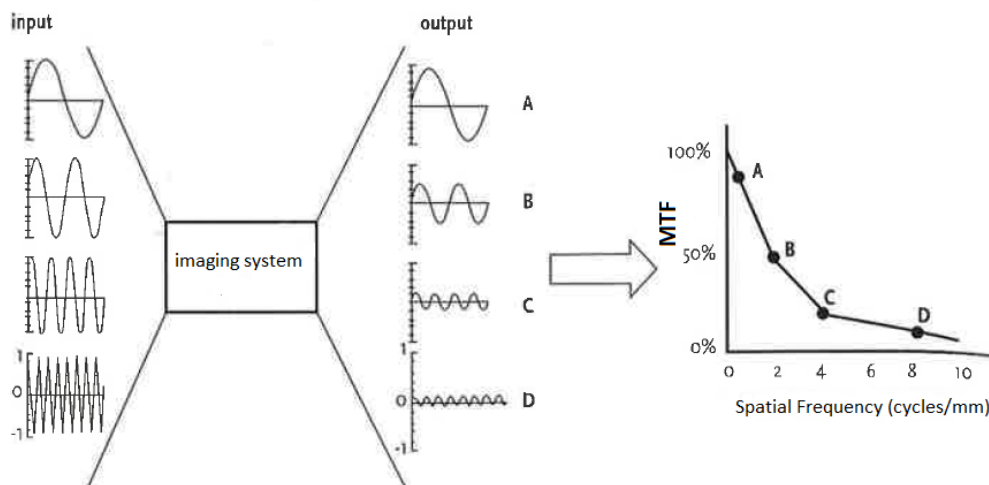


Figure 2-2: Formation of the Modulation Transfer Function [14]

In digital imaging, contrast is mostly described as the Contrast to Noise Ratio (CNR). The CNR is an object size independent measure of the signal level in the investigated tissue in the presence of noise [11]. The contrast is the difference between the average grayscale between a region of interest (ROI) in the investigated tissue (S_t) and the average grayscale in a ROI in the background (S_b). The noise is also calculated in that background ROI (σ_b).

$$CNR = \frac{S_t - S_b}{\sigma_b} [11]$$

All above mentioned image quality parameters are physical measurements to characterize device performance. However, the last stage in the imaging chain is the observer. It is important to include some characterization of the observer's performance in the overall scheme of assessing image quality [12]. This psychophysical approach includes contrast-detail studies, a well-known technique for assessing observer performance of a system. It is a good qualitative way to combine information about the spatial and the contrast resolution. With the

use of a contrast-detail phantom, a test of the observer's perception is possible on a semi objective basis, as the difference between interpretation of the reader and the true distribution of test object can be calculated. After image acquisition of the phantom, the observer(s) must indicate the borderline visibility, resulting in a contrast-detail curve. In the contrast-detail score, a combined effect of image noise, contrast and resolution is included, thereby providing a straightforward measure for the overall performance [15].

In the field of digital radiology and especially digital mammography, the commercially available CDRAD phantom [16] and CDMAM phantom [17] (Artinis Medical Systems BV, Netherlands) are often used to evaluate image quality (*Figure 2-3*). The method originally proposed by the manufacturers for the evaluation of CDRAD/CDMAM images is based on human perception and decision criteria. It involves several observers individually identifying the just visible details (threshold contrast). Human perception makes the evaluation of contrast-detail images a subjective and time consuming task that could be associated with significant inter- and intra-observer errors. Computerized evaluation of the contrast-detail images has the potential to overcome the above limitations [16, 18, 19].

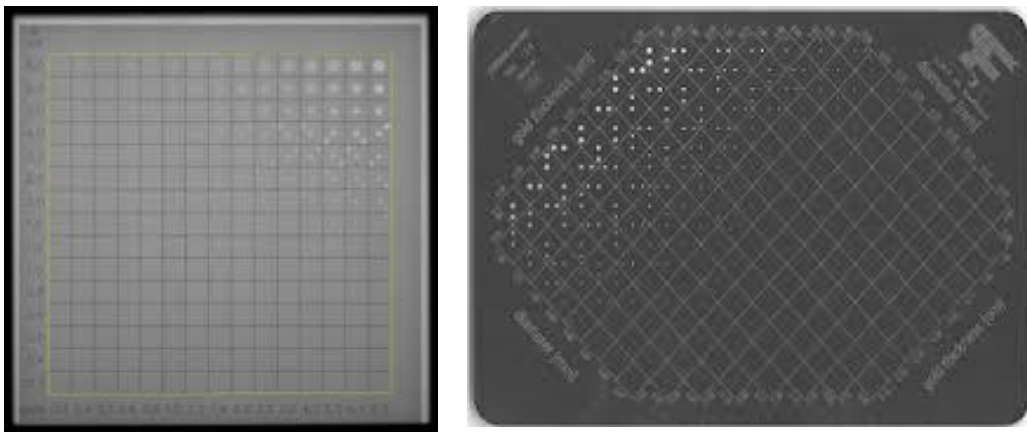


Figure 2-3: Digital acquisition of the CDRAD 2.0 [16] and the CDMAM 3.4 phantom [17].

To evaluate contrast-detail in CT images, a physical-technical image quality phantom can be used like the Catphan@504 phantom (The Phantom laboratory, Salem, New York, USA) (*Figure 2-4*).

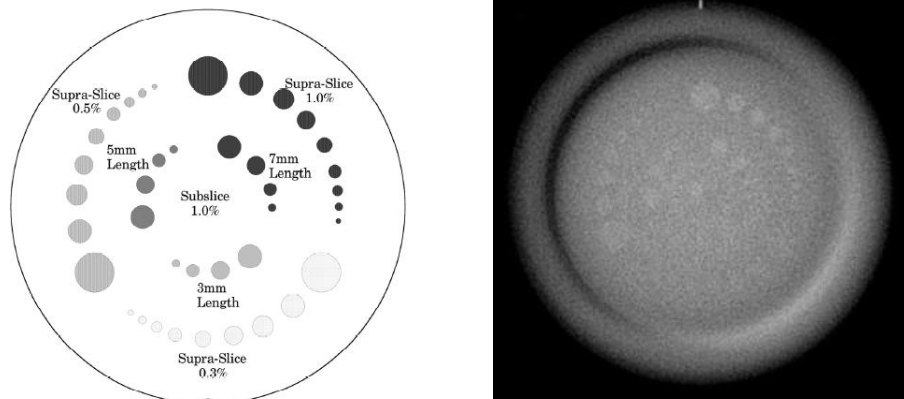


Figure 2-4: Low contrast module of the Catphan 504 phantom [20].

2.3 Patient models

The parameters defined above (NPS, SNR, DQE, PSF, MTF, CNR and contrast-detail) describe mainly the physical-technical image quality as assessed by medical physicists. Clinical image quality, describes the effectiveness with which an image can be used for its intended diagnostic purpose, namely the ability of the image to demonstrate disease and to delineate anatomical structures which are relevant to detection, differential diagnosis and localization [21]. Therefore, image quality should optimally be measured by methods that address clinical performance [22]. For a good approximation of the clinical reality with respect to human anatomy, different study objects can be used. Obviously, patient images are the most realistic to assess clinical image quality. However, due to the variability in anatomy among patients, large numbers of patients should be included for statistical reasons. Consequently, these patient studies are rather difficult to implement since large numbers of patient images must be available. As an alternative, clinical image quality can be assessed in different types of phantoms.

2.3.1 Anthropomorphic phantoms

Anthropomorphic phantoms simulate shape, size and tissue composition of the human body. They have been widely used to provide a representation of the body's anatomy and attenuation characteristics and for radiation dosimetry studies. For dosimetric purposes,

thermoluminescent dosimeters (TLD's) can be inserted in the phantom to register organ doses [23, 24]. Anthropomorphic phantoms are also used to evaluate image quality. Different structures can be superimposed over the phantom on well-defined positions, simulating different pathologies [25-27].

The Alderson Rando phantom is constructed with a natural human skeleton which is cast inside soft tissue simulating material (*Figure 2-5*). Lungs are constructed from lung simulating material and the air space of the head, neck and stem bronchi are duplicated. The phantom is sliced at 2.5 cm intervals. Hole grid patterns are drilled into the sliced sections to enable the insertion of TLD's. It is used to perform radiation dose studies but also for the evaluation of image quality. Since the Rando phantom only contains bone, tissue equivalent material and lung equivalent material, lesions have to be superimposed on the phantom. Another possibility is to digitally superimpose lesions on the image [28, 29].



Figure 2-5: Alderson Rando phantom [30]

The Alderson Lung/Chest Phantom extends from the neck to below the diaphragm (*Figure 2-6*). It is molded about a male skeleton, corresponding to the external body size of a patient of 175 cm and 73.5 kg. Materials are equivalent to natural bone and soft tissues. Animal lungs are selected to match the size of an adult male. Lungs are fixed in the inflated state and are molded to conform to the pleural cavities of the phantom. The pulmonary arteries are injected with a blood equivalent plastic. Custom pathologies are available as wrap around sheets on either or both lungs. All models are available with blood-equivalent pulmonary arteries, or with low or high contrast media added.



Figure 2-6: Alderson Lung/Chest phantom [31].

The CIRS Model 600 Anthropomorphic Torso Phantom is designed to provide an accurate simulation of an average torso (22 cm posterior-anterior thickness) for medical imaging and dosimetry applications (Figure 2-7). The epoxy materials used to fabricate the phantom provide optimal tissue simulation between the diagnostic and therapy X-ray energy range (40 keV to 20 MeV). The Model 600 includes internal organ structures such as the lungs, heart, liver, kidneys, spleen and pancreas. The removable organs enable flexibility in the placement of TLD's, contrast agents, etc. All simulated organs match the tissue density of actual organs and can be clearly visualized.



Figure 2-7: CIRS Model 600 anthropomorphic torso phantom [32].

The Kyoto Kagaku chest phantom N1 Lungman is a multipurpose phantom which is applicable for both plain radiography and CT scanning (*Figure 2-8*). The phantom is an accurate life-size anatomical model of a human torso. The inner components consisting of mediastinum, pulmonary vasculature and an abdomen block are easily detachable, allowing insertion of mimic tumors or other lesions.



Figure 2-8: The Kyoto Kagaku chest phantom [33].

2.3.2 Thiel embalmed cadaver

As an alternative to anthropomorphic phantoms, cadavers can be used (*Figure 2-9*). Cadavers fixed with conventional procedures, by using formalin for conservation, suffer from profound changes of color, strength, fragility and density of organs and tissues. A new embalming technique was established by Prof. Em. Walther Thiel, Anatomisches Institut Karl-Franzens-Universität, Graz, Austria [34]. Hereby, 4-chloro-3-methylenphenol as well as various salts are used for fixation and boric acid is added for disinfection. Furthermore, ethylene glycol is used for preservation of tissue plasticity, while the concentration of formalin is kept to the strict minimum (0.8%) [35]. The first step of the process is perfusion, which takes place shortly after arrival of the body in the anatomy department. The embalming fluids are infused arterially, normally through the femoral or brachial artery. After this, the bodies are submerged in a tank with embalming fluids. The bodies can then be stored in a sealed plastic bag until use, without need for refrigeration. In contrast to standard formalin-embalmed human cadavers, this low-odor technique results in well preserved organs and tissues concerning color, consistency, natural flexibility, density and natural plasticity.



Figure 2-9: Example of a Thiel embalmed cadaver

Solution A:	Boric acid 3% (mono-)Ethylenglycol 30% Ammonium nitrate 20% Potassium nitrate 5% Water Total:	1,9 kg 19 l 12,6 kg 3,2 kg 63,3 l 100 l
Solution B:	(mono-)Ethylenglycol 10% 4-Chlor-3-Methylphenol Total:	18,2 l 1,8 kg 20 l
Embalming fluids:	Solution A Solution B Formalin Sodium sulphate Total:	14,3 l 0,5 l 0,3 l 0,7 kg 15,8 l
Immersion fluids:	(mono-)ethylenglycol 10% Formalin 2% Solution B 2% Boric acid 3% Ammonium nitrate 10% Potassium nitrate 5% Sodium sulphate 7% Water Total:	71,9 l 14,4 l 14,4 l 21,6 kg 71,9 kg 36 kg 50 kg 720 l 1000,2 l

Table 2-1: Composition of Thiel embalming and immersion fluids as used in our faculty

Thiel embalmed cadavers have shown to be suitable as training material for a wide range of surgical procedures such as thyroid surgery [36], laparoscopy [37, 38] and oral surgery [39]. Furthermore, because of the natural flexibility and plasticity of Thiel embalmed cadavers, lung tissue is preserved completely which makes it possible to ventilate the lungs by performing a tracheotomy in combination with balloon ventilation to approximate as nearly as possible the normal patient anatomy [40] (*Figure 2-10*). Consequently, these Thiel embalmed cadavers are an excellent model to investigate the link between clinical and physical-technical image quality.

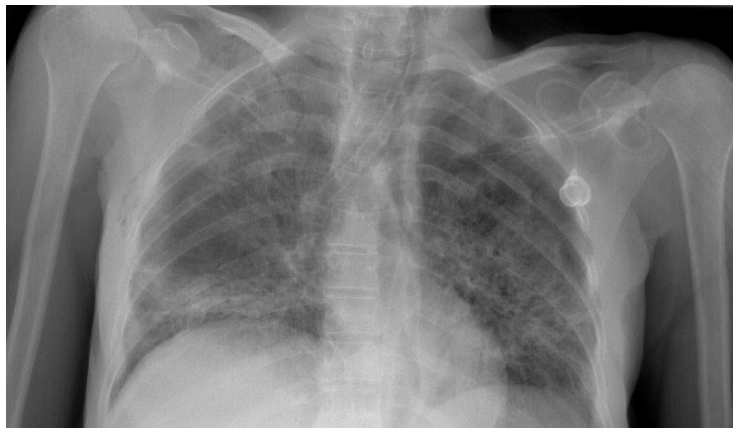


Figure 2-10: Planar chest radiography of a Thiel embalmed cadaver.

2.4 Clinical image quality analysis

To assess the clinical image quality, different study set-ups are possible. The focus of the analysis is either on the normal anatomy of the patient or on the pathological structures present in an image.

2.4.1 Receiver Operator Characteristics (ROC)

The task for the observers in a ROC study is to decide whether a given patient image contains a pathological structure or not. A grading is given according to a scale stating the decision of the observer and also the level of confidence of his or her decision [29]. ROC analysis is based on the fact that the radiologist can adapt to different “critical confidence levels” for calling an image normal or abnormal [41]. Therefore, simple measurement of the sensitivity (the

probability that a patient with an actual disease is determined as having a disease by the observer) and specificity (the probability that a healthy patient is determined as being healthy by the observer) of the diagnosis is not sufficient, but the sensitivity and specificity pairs need to be evaluated for various critical confidence levels. These sensitivity and specificity pairs then define the ROC graph. Although knowledge of the entire ROC curve is required in order to describe the performance of a diagnostic test completely, it would be interesting to summarize the curve by a single number, particularly to draw a conclusion concerning which of two ROC curves is “better”. The most commonly employed univariate summary of a conventional ROC curve is the area under the curve (AUC). It provides a meaningful basis for ranking ROC curves if the curves in question do not cross (*Figure 2-11*).

Real patient images can be used for ROC analysis. However, a large number of patient images and a large number of pathological lesions are required to reach a sufficient statistical power. Furthermore, the true state of the patient from which each image is made must be known to classify an image as normal or abnormal [42]. Unfortunately, the establishment of diagnostic truth in clinical images is sometimes difficult. Furthermore, the lesions need to be subtle so that false positives occur during the image interpretation [29]. To overcome these problems, lesions can be simulated on real patient images [43, 44] Then, knowing the actual truth state in any image is not a problem. However, this still requires a large number of patient images. Though, ROC studies are not necessarily based on actual patient images. Instead, images of a suitable phantom in which a relevant (simulated) signal is added, can be used [22].

The ROC method is not applicable when the decision task involves more than a simple determination of whether the patient shows a disease symptom or not. In addition to detecting an abnormal condition, the radiologist often needs to locate specific image regions that are suspicious for disease. The additional location information cannot be used by ROC analysis and this neglect of location information may lead to a loss of statistical power, so that actual differences between modalities may be overlooked [45]. For example, if an observer misses the single true lesion in an image but erroneously identifies another location as containing a lesion, the observer makes two mistakes: a false negative and a false positive. However, the two mistakes effectively cancel out each other, and on the case level the observer is scored with a true positive.

Another disadvantage of ROC and ROC related methods, is that the results are only valid for the type of pathological structure that is used in the study [42]. The results cannot be

extrapolated to hold for other types of lesions. To study the performance of a radiographic system in a clinical environment where the variation of the lesions is large, ROC studies with images containing the whole spectrum of disease-indicating symptoms that are present in the actual patient population have to be set up [29]. This can easily lead to unrealistically large number of images that have to be observed in the trial to achieve the desired statistical precision.

To overcome the limitations of ROC analysis, several other methods have been developed. Bunch et al. [46] proposed the free response ROC, FROC, in which location information is considered. In a FROC experiment, several lesions may be used in each image, localization of each lesion must be performed, and a rating reflecting the observer's certainty of the lesion and its location must be given. The FROC curve gives the fraction of true positive localized lesions as a function of the mean number of false-positive answers per image.

Alternative FROC (AFROC) analysis is an alternative way of analyzing FROC data. Instead of using the mean number of false positives per image, Chakraborty and Winter [47] suggested to use the probability that an image produces one or more false positive responses. The AFROC curve is constructed by plotting the true positive fraction (TPF) of detected and localized lesions as a function of the probability of at least one false positive per image.

With the Free-response forced error (FFE) method, the observer is asked to rank the test structures in decreasing order of confidence until an error is made [47]. The average fraction of correct findings over a number of images before a false positive location is indicated, is equal to the area under the AFROC curve.

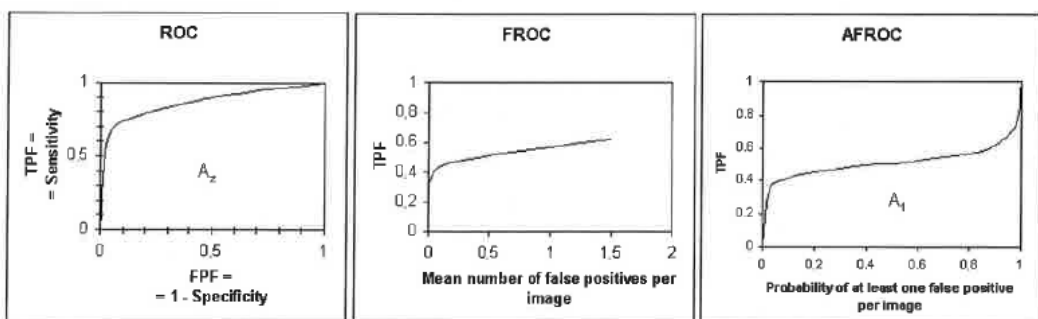


Figure 2-11: Examples of curves resulting from ROC, FROC and AFROC [48].

2.4.2 Visual Grading Analysis (VGA)

In visual grading analysis, contrary to the ROC related methods, it is the normal anatomy, present in almost every patient image, which is used for the evaluation of image quality. The task for the observer is to decide whether the criteria related to the normal anatomy are fulfilled or not in an image. The evaluated structures can be selected so that they represent different characteristics that can be found for typical pathological structures [29], or can be selected from the criteria defined by the Commission of the European Communities (CEC). The CEC has presented a list of image criteria in the European guidelines on Quality Criteria for Diagnostic Radiographic Images [49] and in the European guidelines on Quality Criteria for Computed Tomography [50]. The validity of such studies can assumed to be high if the anatomical structures are selected based on their clinical relevance. VGA can be performed in two major ways: with relative grading (using one or several images as references) or with absolute grading (using no references) [51].

A VGA score (VGAS) can be calculated as [52]:

$$VGAS = \frac{\sum_{s=1}^S \sum_{i=1}^I G_{s,i}}{S * I}$$

where $G_{s,i}$ is the rating for a particular structure (s) and image (i). S and I are the number of structures and images, respectively.

The results of a VGA study are more general than those from a ROC or ROC related study in the sense that anatomical structures of different size, shapes and textures are used for the evaluation. An important disadvantage of VGA is that it does not include the clinical task, i.e. to differentiate between normal and pathological cases.

2.5 Observer Models

An observer can be either a human or an algorithm (model observer). The use of human observers is labor intensive. Humans are also subject to effects like fatigue or reader learning that introduce bias [53]. Computer-model observers are algorithms that attempt to predict human visual performance in noisy images and might represent the desired metric of image

quality [54]. Development of models to predict human visual signal detection in noise goes back to work by Rose [55, 56] who studied the detectability of a flat-topped disk embedded in white noise. In the last years, many studies have concentrated on finding a model observer that can predict human performance across many types of synthetic backgrounds and real medical-image backgrounds [53, 57-63]. The hope is that model observers will become common metrics of task-based image quality for evaluation of medical image quality as well as optimization of imaging systems. In order to use model observers, a number of key components must be defined.

Firstly, the investigator needs to choose the visual task which the model observer has to evaluate or optimize. Medical imaging tasks can be categorized as either classification or estimation tasks [63, 64]. A commonly used classification task is the multiple alternative forced choice task (MAFC), where the lesion might appear in one of M specified locations. The task of the observer is to indicate which of the M locations contains the target [53, 65, 66]. Other tasks, such as a yes/no task with no location uncertainty, can also easily be used for the model observer [66]. An estimation task involves the quantification of one or more parameters that describe the object. The parameter might be e.g. the amount of flow in a vessel or the size, location, or activity of a tumor [54].

Secondly, to set up a model observer study, decisions have to be made about the type of backgrounds and signals that will be used to evaluate a particular imaging modality and imaged anatomy. One approach is to use computer simulated backgrounds that visually appear similar to the real image backgrounds for the particular imaging modality, with the hope that the model observer results with the synthetic backgrounds can be generalized to the real background [65]. A different approach is to use real anatomic backgrounds from patient data rather than computer simulated backgrounds [57, 58, 65]. This latter method requires a large database of patient images so that the model observer results can be generalized to the population of patient images. The advantage of the computer simulated backgrounds is that the statistical properties of the backgrounds are known and satisfy many statistical properties that make calculation of model performance simpler [62]. The advantage of real anatomic backgrounds is clearly their clinical realism. The disadvantage is that often real anatomic backgrounds do not satisfy certain statistical properties and therefore some of the standard techniques for calculating model performance do not always generalize to these real background images [62].

The next step is to select a signal to be detected in the visual task. Signals almost always are computer generated and often are designed to mimic a realistic feature of interest in the medical images [57, 58].

Thirdly, a model observer or a set of model observers to perform the visual task must be chosen from the large variety of available models. Early efforts on model observers concentrated on the ideal Bayesian observer [67, 68]. The Bayesian ideal observer provides a quantitative measure of the diagnostic performance of an imaging system. This observer is optimal among all observers, either human or model, and sets an absolute upper bound for observer performance in classification tasks [69]. However, the Bayesian ideal observer requires the full description of the statistical properties of the data to optimally perform a classification task, but such information is often unknown for complex, realistic backgrounds found in clinical applications [70].

Most of the model observers currently used for visual detection in noisy images are linear models. Models can be classified based on the prior knowledge they use. For example, some models use information about the signal; others use information about the signal and the background statistics. Given that these models are attempting to predict human performance, many of them include information-processing constraints intended to reflect properties of the human visual system derived from psychophysical or physiological findings [54].

Fourth, a method to calculate model observer performance in the selected visual task must be determined. This figure of merit quantifies the ability of the observer to use the image data to perform the relevant task and is an objective measure of image quality [63].

And finally, if human performance in the same task has been measured then a method for comparing model and human performance is required [54]. Multiple model observers have been proposed to objectively evaluate image quality and optimize system design. For any model observer, studies are required to demonstrate its correlation with human observer performance before it can be used clinically [70]. Given the substantial difference between imaging modalities concerning signal and noise properties, dedicated studies need to be performed for each particular modality [71]. A typical problem is that performance of the model observers is typically better than that of human performance [57]. A commonly used method to compare model and human performance is to add additional sources of degradation to the model observer so that the model observer performs at human levels [54].

This can be achieved by including internal noise in the model. Internal noise is a known component of human inefficiency in perceptual tasks and arises from fluctuations in neural firing, intrinsic stimulus variability, receptor sampling errors and loss of information during neural transmission [72]. Internal noise can explain the fact that human observers make different decisions on the same set of images.

References

1. UNSCEAR, *Vol I: Sources and effects of ionizing radiation, Annex A: Medical radiation exposures. Report to the General Assembly, United Nations, New York.* 2010.
2. Dose Datamed, *DDM2 Project Report Part 1: European Population Dose: Study on European Population Doses from Medical Exposure.* 2014.
3. UNSCEAR, *Sources and effect of ionizing radiation - Vol I: Sources. Report to the General assembly, United Nations, New York.* 2000.
4. RIZIV, *Data received by the department of medical care of the RIZIV. Thanks to Engels H.* 2014.
5. ICRP, *Publication 73: Radiological Protection and Safety in Medicine.* Annals of the ICRP volume 26. 1996, Oxford: Pergamon.
6. The Council of the European Union, *Council Directive 97/43/Euratom. Official Journal of the European Communities.* 1997.
7. ICRP, *Publication 60: Recommendations of the International Commission on Radiological Protection.* Annals of the ICRP volume 21. 1991, Oxford: Pergamon.
8. Federal Agency of Nuclear Control (FANC), *Data received by the FANC.* 2014.
9. Samei, E., W.R. Eyler, and L. Baron, *Handbook of Medical Imaging. Vol 1, Physics and Psychophysics: Effects of anatomical structure on signal detection.* 2000, Bellingham, Washington: SPIE Press.
10. Samei, E., *Advances in Digital Radiography: RSNA Categorical Course in Diagnostic Radiology Physics.* 2003: p. 37-47.
11. Bushberg, J., et al., *The essential physics of medical imaging.* 2nd ed. 2002, Philadelphia, PA: Lippincott Williams & Wilkins.
12. Dobbins, J.T., *Handbook of Medical Imaging. Vol 1, Physics and Psychophysics: Image quality metric for digital system.* 2000, Bellingham, Washington: SPIE Press.
13. Goodenough, D.J., *Handbook of Medical Imaging. Vol 1, Physics and Psychophysics: Tomographic imaging.* 2000, Bellingham, Washington: SPIE Press.
14. Hensen, J., et al., *Radiologie: Techniek en onderzoek.* 2011, Amsterdam: Reed Business.
15. Thijssen, M., K. Bijkerk, and R. van der Burgth, *Manual Contrast-Detail Phantom CDRAD type 2.0. Project Quality Assurance in Radiology, Department of Radiology, University Hospital Nijmegen, St. Radboud, The Netherlands.* 1998.
16. Artinis Medical Systems, *Manual CDRAD 2.0 Phantom & analyser software version 2.1.*
17. Thijssen, M. and K.R. Bijkerk, *Manual Contrast-detail phantom Artinis CDMAM type 3.4, Artinis Medical Systems. Department of Radiology, University Medical Center Nijmegen, The Netherlands.*
18. Pascoal, A., et al., *Evaluation of a software package for automated quality assessment of contrast detail images - comparison with subjective visual assessment.* Physics in Medicine and Biology, 2005. **50**(23): p. 5743-5757.
19. Christodoulou, E., et al., *Evaluation of the contrast-detail response of digital radiographic systems using the CDRAD contrast-detail phantom with the CDRAD Analyser software.* Medical Physics, 2006. **33**(6): p. 2008-2008.
20. The phantom laboratory, *Catphan 504 Manual.* 2013.

21. Martin, C.J., P.F. Sharp, and D.G. Sutton, *Measurement of image quality in diagnostic radiology*. Applied Radiation and Isotopes, 1999. **50**(1): p. 21-38.
22. Tapiovaara, M.J., *Review of relationships between physical measurements and user evaluation of image quality*. Radiation Protection Dosimetry, 2008. **129**(1-3): p. 244-248.
23. Loubele, M., et al., *Comparison between effective radiation dose of CBCT and MSCT scanners for dentomaxillofacial applications*. European Journal of Radiology, 2009. **71**(3): p. 461-468.
24. Svenson, B., B. Sjöholm, and B. Jonsson, *Reduction of absorbed doses to the thyroid gland in orthodontic treatment planning by reducing the area of irradiation*. Swedish Dental Journal, 2004. **28**(3): p. 137-147.
25. Oschatz, E., et al., *Comparison of liquid crystal versus cathode ray tube display for the detection of simulated chest lesions*. European Radiology, 2005. **15**(7): p. 1472-1476.
26. Uffmann, M., et al., *Computed radiography and direct radiography - Influence of acquisition dose on the detection of simulated lung lesions*. Investigative Radiology, 2005. **40**(5): p. 249-256.
27. Metz, S., et al., *Chest radiography with a digital flat-panel detector: Experimental receiver operating characteristic analysis*. Radiology, 2005. **234**(3): p. 776-784.
28. Cho, H.M., et al., *ROC Analysis of Storage Phosphor Radiography and Digital Selenium Radiography at Various Tube Voltages in Simulated Chest Lesions*, in 2008 IEEE Nuclear Science Symposium and Medical Imaging Conference. 2009, IEEE: New York. p. 4825-4829.
29. Tingberg, A., et al., *Evaluation of image quality of lumbar spine images: A comparison between FFE and VGA*. Radiation Protection Dosimetry, 2005. **114**(1-3): p. 53-61.
30. The phantom laboratory, *Rando phantom manual*. 2012.
31. Radiology support devices website. Available from http://www.rsdphantoms.com/rd_lung.htm.
32. CIRS, *Manual 3D sectional torso phantom*. 2014.
33. Kyoto Kagaku, *Manual chest phantom N1 lungman*.
34. Thiel, W., *Ergänzung für die Konservierung ganze Leichen nach W. Thiel*. Annals of Anatomy, 2002. **184**: p. 267-269.
35. Thiel, W., *Die Konservierung ganzer Leichen in natürlichen Farben*. Annals of Anatomy, 1992. **174**: p. 185-195.
36. Eisma, R., et al., *A comparison of Thiel and formalin embalmed cadavers for thyroid surgery training*. Surgeon, 2011. **9**(3): p. 142-6.
37. Kerckaert, I., et al., *Endogent: Centre for Anatomy and Invasive Techniques*. International Journal of Experimental and Clinical Anatomy, 2008. **2**: p. 28-33.
38. Giger, U., et al., *Laparoscopic training on Thiel human cadavers: a model to teach advanced laparoscopic procedures*. Surg Endosc, 2008. **22**(4): p. 901-6.
39. Holzle, F., et al., *Thiel embalming technique: a valuable method for teaching oral surgery and implantology*. Clin Implant Dent Relat Res, 2012. **14**(1): p. 121-6.
40. De Crop, A., et al., *Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study*. Radiology, 2012. **262**(1): p. 298-304.
41. Bath, M., *Evaluating imaging systems: practical applications*. Radiat Prot Dosimetry, 2010. **139**(1-3): p. 26-36.
42. Metz, C., *Handbook of Medical Imaging. Vol 1, Physics and Psychophysics: Fundamental ROC Analysis*. 2000, Bellingham, Washington: SPIE Press.

43. Solomon, J. and E. Samei, *A generic framework to simulate realistic lung, liver and renal pathologies in CT imaging*. Physics in Medicine and Biology, 2014. **59**(21): p. 6637-6657.
44. Yano, Y., et al., *Detectability of simulated pulmonary nodules on chest radiographs: Comparison between irradiation side sampling indirect flat-panel detector and computed radiography*. European Journal of Radiology, 2013. **82**(11): p. 2050-2054.
45. Chakraborty, D., *Statistical power in observer-performance studies: Comparison of the receiver operating characteristic and free-response methods in tasks involving localization*. Academic Radiology, 2002. **9**(2): p. 147-156.
46. Bunch, P.C., et al., *FREE-RESPONSE APPROACH TO THE MEASUREMENT AND CHARACTERIZATION OF RADIOGRAPHIC-OBSERVER PERFORMANCE*. Journal of Applied Photographic Engineering, 1978. **4**(4): p. 166-171.
47. Chakraborty, D.P. and L.H.L. Winter, *FREE-RESPONSE METHODOLOGY - ALTERNATE ANALYSIS AND A NEW OBSERVER-PERFORMANCE EXPERIMENT*. Radiology, 1990. **174**(3): p. 873-881.
48. Tingberg, A., *Quantifying the quality of medical X-ray images: An evaluation based on normal anatomy for lumbar spine and chest radiography*, in Department of Radiation Physics, Malmö. 2000, Lund University, Sweden.
49. Commission of the European Communities, *European guidelines on quality criteria for diagnostic radiographic images (EUR 16260)*. 1996.
50. Commission of the European Communities, *European guidelines on quality criteria for computed tomography (EUR 16262)*.
51. Mansson, L.G., *Methods for the evaluation of image quality: A review*. Radiation Protection Dosimetry, 2000. **90**(1-2): p. 89-99.
52. Sund, P., et al., *Comparison of visual grading analysis and determination of detective quantum efficiency for evaluation system performance in digital chest radiography*. European Radiology, 2004. **14**: p. 48-58.
53. Vaishnav, J.Y., et al., *Objective assessment of image quality and dose reduction in CT iterative reconstruction*. Medical Physics, 2014. **41**(7).
54. Eckstein, M.P., C. Abbey, and F. Bochud, *Handbook of Medical Imaging. Vol 1, Physics and Psychophysics: A practical guide to model observers for visual detection in synthetic and natural noisy images*. 2000, Bellingham, Washington: SPIE Press.
55. Rose, A., *The sensitivity performance of the human eye on an absolute scale*. J Opt Soc Am, 1948. **38**(2): p. 196-208.
56. Burgess, A.E., *The Rose model, revisited*. Journal of the Optical Society of America a-Optics Image Science and Vision, 1999. **16**(3): p. 633-646.
57. Zhang, Y., B.T. Pham, and M.P. Eckstein, *Evaluation of internal noise methods for Hotelling observer models*. Medical Physics, 2007. **34**(8): p. 3312-3322.
58. Zhang, Y., B. Pham, and M.P. Eckstein, *Evaluation of JPEG 2000 encoder options: Human and model observer detection of variable signals in X-ray coronary angiograms*. Ieee Transactions on Medical Imaging, 2004. **23**(5): p. 613-632.
59. Veldkamp, W.J.H., et al., *A Technique for Simulating the Effect of Dose Reduction on Image Quality in Digital Chest Radiography*. Journal of Digital Imaging, 2009. **22**(2): p. 114-125.
60. Srinivas, Y. and D.L. Wilson, *Quantitative image quality evaluation of pixel-binning in a flat-panel detector for x-ray fluoroscopy*. Medical Physics, 2004. **31**(1): p. 131-141.

61. Diaz, I., et al., *Development of model observers applied to 3D breast tomosynthesis microcalcifications and masses*. Medical Imaging 2011: Image Perception, Observer Performance, and Technology Assessment, 2011. **7966**.
62. Bochud, F.O., C.K. Abbey, and M.P. Eckstein, *Visual signal detection in structured backgrounds. III. Calculation of figures of merit for model observers in statistically nonstationary backgrounds*. Journal of the Optical Society of America a-Optics Image Science and Vision, 2000. **17**(2): p. 193-205.
63. Tseng, H.W., et al., *Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers*. Med Phys, 2014. **41**(7): p. 071910.
64. He, X. and S. Park, *Model observers in medical imaging research*. Theranostics, 2013. **3**(10): p. 774-86.
65. Castella, C., et al., *Human linear template with mammographic backgrounds estimated with a genetic algorithm*. Journal of the Optical Society of America a-Optics Image Science and Vision, 2007. **24**(12): p. B1-B12.
66. Eckstein, M.P., C.K. Abbey, and F.O. Bochud, *Visual signal detection in structured backgrounds. IV. Figures of merit for model performance in multiple-alternative forced-choice detection tasks with correlated responses*. Journal of the Optical Society of America a-Optics Image Science and Vision, 2000. **17**(2): p. 206-217.
67. Burgess, A.E., et al., *EFFICIENCY OF HUMAN VISUAL SIGNAL DISCRIMINATION*. Science, 1981. **214**(4516): p. 93-94.
68. Wagner, R.F. and D.G. Brown, *Unified Snr Analysis of Medical Imaging-Systems*. Physics in Medicine and Biology, 1985. **30**(6): p. 489-518.
69. Park, S. and E. Clarkson, *Efficient estimation of ideal-observer performance in classification tasks involving high-dimensional complex backgrounds*. Journal of the Optical Society of America a-Optics Image Science and Vision, 2009. **26**(11): p. B59-B71.
70. Barrett, H.H., et al., *Model Observers for Assessment of Image Quality*. Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(21): p. 9758-9765.
71. Leng, S., et al., *Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain*. Medical Physics, 2013. **40**(8).
72. Lu, Z.L. and B.A. Doshier, *Characterizing human perceptual inefficiencies with equivalent internal noise*. Journal of the Optical Society of America a-Optics Image Science and Vision, 1999. **16**(3): p. 764-778.

Part II:

Aim and outline of the thesis

3

Aim of the thesis

Image quality analysis represents a crucial component in the radiation dose optimization process. Physical-technical image quality parameters focus more on the performance of the detector and do not address the clinical performance of the imaging modality. Clinical image quality describes the effectiveness at which an image can be used for its intended purpose. However, the link between clinical and physical-technical image quality is still an open question. The lack of knowledge on this correlation has a significant impact on the use of physical-technical image quality parameters in quality assurance protocols, since the impact of changes in physical-technical quality on the clinical image quality can't be quantified. Up till now, only the European protocol for quality assurance of screening mammography equipment, describes the use of MTF, DQE, NPS, noise evaluation and contrast-detail analysis [1]. Therefore, the first aim of this thesis was to investigate the correlation between clinical and physical-technical image quality in digital radiography, more specifically, chest radiography.

The number of CT examinations has increased rapidly over the last few years, resulting in a substantial increase in radiation dose of the population in the Western world [2]. It has been estimated that these CT examinations may be responsible for approximately 2% of all incident cancer cases in the United States [3]. Consequently, a lot of efforts have been made over the last decade to reduce the radiation dose for the patient by introducing new techniques such as automatic tube current modulation and adaptive collimation [4-6]. Over the past few years, iterative reconstruction techniques are increasingly gaining acceptance [7-13]. If dose

reduction techniques are implemented, the impact on the image quality has to be investigated.

Similar to digital radiology, the link between clinical and physical-technical image quality is not worked out in CT. Particularly for noise, this can be problematic, since noise measurements in a uniform phantom don't account for the complex relationship between anatomical variability and image quality [14].

The second aim of this thesis was to evaluate the correlation between clinical and physical-technical image quality applied to different strengths of iterative reconstruction in chest CT images and to determine the potential dose reduction of iterative reconstruction compared to conventional filtered back projection based on different clinical and physical-technical image quality parameters.

CT plays a very important role in both diagnosis and treatment of head and neck cancers. It has become the standard method for identifying cancerous masses, lymph node metastases and inflammatory processes in head and neck imaging [15-17]. The image quality and associated suitability for accurate diagnoses can be significantly reduced by dental artifacts. The factors that may contribute to or diminish the artefacts include metallic hardware composition and orientation [18-20], peak voltage [21, 22], slice thickness [21], reconstruction algorithm [18, 23], extended CT scale [24]. Alternatively, artifacts can be suppressed during reconstruction. Metal artefact reduction algorithms have been developed for this purpose.

The third aim was to investigate the efficiency of different metal artefact reduction tools such as kVp, iterative reconstruction, virtual monochromatic imaging (keV) and metal artefact reduction software (MARs).

4

Outline of the thesis

Chest radiography is one of the most frequently performed diagnostic X-ray examinations [25]. The frequent use and diagnostic importance of chest X-ray makes it an interesting topic for optimization of image quality and patient dose. Therefore, the correlation between clinical and physical-technical image quality was assessed for digital chest radiography. Clinical image quality of chest radiographs was assessed by means of three human cadavers, conserved using the Thiel embalming technique [26]. Physical-technical image quality was assessed using the CDRAD 2.0 contrast-detail phantom. Four experienced radiologists assessed the image quality, using a visual grading analysis (VGA) technique based on the European Quality Criteria for Chest Radiology [27]. The CDRAD images were scored manually and automatically using dedicated software (CDRAD analyser 2.1) [28], both resulting in an Image Quality Figure inverse (IQF_{inv}) value. The results of this first study are presented in a first paper (see chapter 5) entitled CORRELATION OF CONTRAST-DETAIL ANALYSIS AND CLINICAL IMAGE QUALITY ASSESSMENT IN CHEST RADIOGRAPHY WITH A HUMAN CADAVER STUDY.

As the first study showed an excellent correlation between clinical and physical-technical image quality in digital chest radiography, the correlation in chest CT was investigated in the second study. Therefore the potential dose reduction of iterative reconstruction compared to conventional filtered back projection based on different clinical and physical-technical image quality parameters was assessed. Clinical image quality was assessed using three Thiel embalmed human cadavers. Four experienced radiologists assessed the image quality, using a visual grading analysis (VGA) technique based on the European Quality Criteria for Chest

Tomography [29]. A Catphan phantom was used to assess physical-technical image quality parameters such as noise, contrast-detail and contrast-to-noise ratio (CNR). The results from this study are presented in the second paper (see chapter 6) entitled CORRELATION OF CLINICAL AND PHYSICAL-TECHNICAL IMAGE QUALITY IN CHEST CT: A HUMAN CADAVER STUDY APPLIED ON ITERATIVE RECONSTRUCTION.

A variety of tools is available to diminish dental metallic artifacts in head and neck CT imaging. However, up till now, no comparison of the effect of different tools was made. Therefore, the diminishing effect on artifacts of the following techniques was studied: altering kVp, application of model based iterative reconstruction (Veo), use of virtual monochromatic images and metal artifact reduction software (MARs). Clinical image quality was assessed using a Thiel embalmed cadaver. Four experienced radiologists assessed the image quality, using a visual grading analysis (VGA) technique. Physical-technical image quality was assessed in a uniform polymethylmethacrylate (PMMA) cylindrical phantom and in the Catphan phantom. The results from this study are presented in the third paper (see chapter 7) entitled: ANALYSIS OF METAL ARTIFACT REDUCTION TOOLS FOR DENTAL HARDWARE IN CT SCANS OF THE ORAL CAVITY: kVp, ITERATIVE RECONSTRUCTION, DUAL ENERGY CT, METAL ARTIFACT REDUCTION SOFTWARE: DOES IT MAKE A DIFFERENCE?

References

1. EUREF, E.R.O.f.Q.A.B.S.a.D.S., *European guidelines for quality assurance in breast cancer screening and diagnosis; fourth edition*. 2013.
2. ICRP, *Publication 87: Managing patient dose in computed tomography*. 2000.
3. Brenner, D.J. and E.J. Hall, *Computed Tomography - An Increasing Source of Radiation Exposure*. The New England Journal of Medicine, 2007. **357**: p. 2277-2284.
4. Lee, T.-Y. and R.K. Chhem, *Impact of new technologies on dose reduction in CT*. European Journal of Radiology, 2010. **76**: p. 28-35.
5. Kalra, M.K., et al., *Strategies for CT radiation dose optimization*. Radiology, 2004. **230**(3): p. 619-28.
6. Deak, P.D., et al., *Effects of Adaptive Section Collimation on Patient Radiation Dose in Multisection Spiral CT*. Radiology, 2009. **252**(1): p. 140-147.
7. Baker, M.E., et al., *Contrast-to-Noise Ratio and Low-Contrast Object Resolution on Full- and Low-Dose MDCT: SAFIRE Versus Filtered Back Projection in a Low-Contrast Object Phantom and in the Liver*. American Journal of Roentgenology, 2012. **199**(1): p. 8-18.
8. Ghetti, C., O. Ortenzia, and G. Serreli, *CT iterative reconstruction in image space: A phantom study*. Physica Medica-European Journal of Medical Physics, 2012. **28**(2): p. 161-165.
9. Ghetti, C., et al., *Physical characterization of a new CT iterative reconstruction method operating in sinogram space*. Journal of Applied Clinical Medical Physics, 2013. **14**(4): p. 263-271.
10. Khawaja, R.D.A., et al., *Ultra-low dose abdominal MDCT: Using a knowledge-based Iterative Model Reconstruction technique for substantial dose reduction in a prospective clinical study*. European Journal of Radiology, 2015. **84**(1): p. 2-10.
11. Kriegshauser, J.S., et al., *Feasibility of ultra-low radiation dose reduction for renal stone CT using model-based iterative reconstruction: prospective pilot study*. Clinical Imaging, 2015. **39**(1): p. 99-103.
12. Bae, S., et al., *Effects of adaptive statistical iterative reconstruction on radiation dose reduction and diagnostic accuracy of pediatric abdominal CT*. Pediatric Radiology, 2014. **44**(12): p. 1541-1547.
13. Chen, B.Y., et al., *Evaluating iterative reconstruction performance in computed tomography*. Medical Physics, 2014. **41**(12): p. 11.
14. Solomon, J. and E. Samei, *Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE*. Medical Physics, 2014. **41**(9): p. 12.
15. Geets, X., et al., *Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: Comparison between CT-scan and MRI*. Radiotherapy and Oncology, 2005. **77**(1): p. 25-31.
16. Gonzalez-Beicos, A. and D. Nunez, *Imaging of Acute Head and Neck Infections*. Radiologic Clinics of North America, 2012. **50**(1): p. 73-+.
17. Cho, J.J.W., et al., *Impact of Newer Generation Multidetector Computed Tomography on the Diagnosis of Abscesses in the Head and Neck*. Journal of Otolaryngology-Head & Neck Surgery, 2011. **40**(4): p. 6.

18. Lee, M.J., et al., *Overcoming artifacts from metallic orthopedic implants at high-field-strength MR imaging and multidetector CT*. Radiographics, 2007. **27**(3): p. 791-803.
19. Fiala, T.G.S., R.A. Novelline, and M.J. Yaremchuk, *Comparison of Ct Imaging Artifacts from Craniomaxillofacial Internal-Fixation Devices*. Plastic and Reconstructive Surgery, 1993. **92**(7): p. 1227-1232.
20. Haramati, N., et al., *Ct Scans through Metal Scanning Technique Versus Hardware Composition*. Computerized Medical Imaging and Graphics, 1994. **18**(6): p. 429-434.
21. Moon, S.G., et al., *Metal Artifact Reduction by the Alteration of Technical Factors in Multidetector Computed Tomography: A 3-Dimensional Quantitative Assessment*. Journal of Computer Assisted Tomography, 2008. **32**(4): p. 630-633.
22. Lee, I.S., et al., *A pragmatic protocol for reduction in the metal artifact and radiation dose in multislice computed tomography of the spine: Cadaveric evaluation after cervical pedicle screw placement*. Journal of Computer Assisted Tomography, 2007. **31**(4): p. 635-641.
23. Stradiotti, P., et al., *Metal-related artifacts in instrumented spine. Techniques for reducing artifacts in CT and MRI: state of the art*. European Spine Journal, 2009. **18**: p. S102-S108.
24. Link, T.M., et al., *CT of metal implants: Reduction of artifacts using an extended CT scale technique*. Journal of Computer Assisted Tomography, 2000. **24**(1): p. 165-172.
25. Dose Datamed, *DDM2 Project Report Part 1: European Population Dose: Study on European Population Doses from Medical Exposure*. 2014.
26. Thiel, W., *Die Konservierung ganzer Leichen in natürlichen Farben*. Annals of Anatomy, 1992. **174**: p. 185-195.
27. Commission of the European Communities, *European guidelines on quality criteria for diagnostic radiographic images (EUR 16260)*. 1996.
28. Artinis Medical Systems, *Manual CDRAD 2.0 Phantom & analyser software version 2.1*.
29. Commission of the European Communities, *European guidelines on quality criteria for computed tomography (EUR 16262)*.

Part III:

Original research

Correlation of Contrast-Detail Analysis and Clinical Image Quality Assessment in Chest Radiography with a Human Cadaver Study

5.1 Background

5.1.1 Digital radiography

Projection radiography has evolved in many radiology departments from a technique dominated by screen-film combinations to a digital technique based on flat panel detectors (Direct Radiography, DR) or imaging plates containing a phosphor layer (Computed Radiography, CR). Workflow has been improved and simplified because of the instant image display, eradicating the time-consuming development of films, storage and archiving of images.

The dynamic range of a detector describes the range over which suitable imaging performance may be obtained (*Figure 5-1*). This is defined basically by the ratio of the maximum to minimum detector entrance doses [1]. This response is referred to as the characteristic curve of the device, and in the case of photographic film, has the well-known sigmoid shape of the Hurter-Drillfield (HD) curve. Because of this, underexposed films appear light, and overexposed films appear dark. Digital systems, on the other hand, have a linear characteristic

curve [2]. Consequently, DR and CR detectors have the benefits of wide dose latitude and improved information recording across different tissue densities.

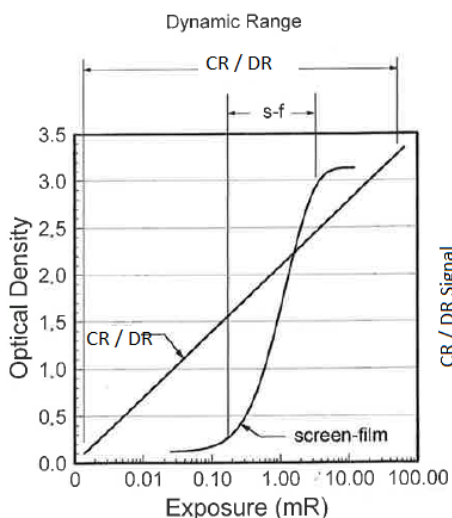


Figure 5-1: Dynamic range of screen-film and digital (CR and DR) detectors

Digital radiography separates acquisition, processing and display, which enables a radiographer to produce an image that has acceptable diagnostic quality, but could be underexposed or overexposed [3]. Adjustments to compensate for exposure technique errors can be made at the time of display. The ability to correct over- or underexposed images in CR and DR radiographic systems by simply changing the window and/or level, has led to the tendency to be less concerned about the exposure technique and creates the opportunity to use more radiation than necessary. Moreover, in digital radiography, the computer automatically adjusts an image that is overexposed to ensure that the image is of diagnostic quality. This automatic adjustment combined with the separation of image acquisition and display can contribute to increased patient exposure [3]. An excessive exposure to a patient during a digital radiography examination does not affect image quality, except at extremely high levels of exposure. In fact, the decreased image noise that results from additional exposure can lead to a corresponding decrease in complaints from radiologists regarding image quality.

Overall, digital radiography may have the tendency to lead to increased radiation dose, therefore, optimization in digital radiography is of critical importance.

5.1.2 Thorax radiography

Chest radiography is one of the most frequently performed diagnostic radiographic examinations. In 2011, the frequency of chest radiography in Belgium was 230 per 1000 of population [4]. In the same period, the average amount of chest radiographs over 36 European countries was 194 per 1000 of population [4] (*Figure 5-2*). In 2012 and 2013, 31% of all radiography examinations in Belgium were chest radiographs [5].

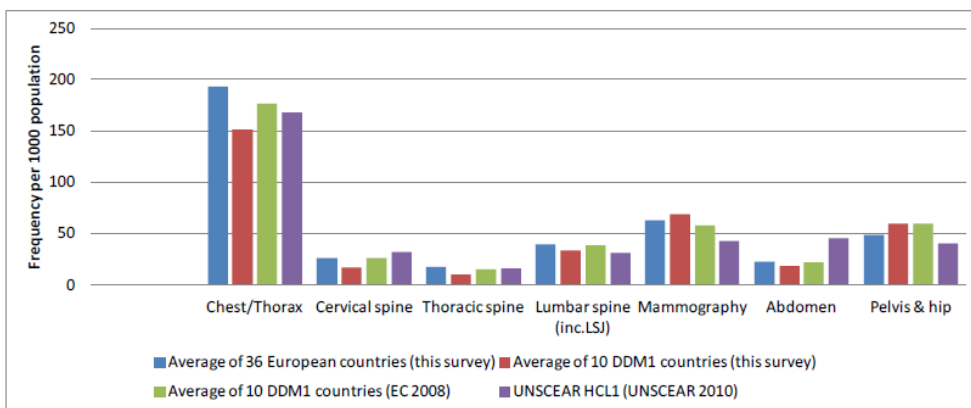


Figure 5-2: Average frequencies per 1000 of population for plain radiography examinations [4]

Although the frequency of chest radiography is high, the contribution to the collective dose is rather low. The mean effective dose estimated in European countries for plain chest radiography is 0.10 mSv, in Belgium this is 0.09 mSv [4] (*Figure 5-3*).

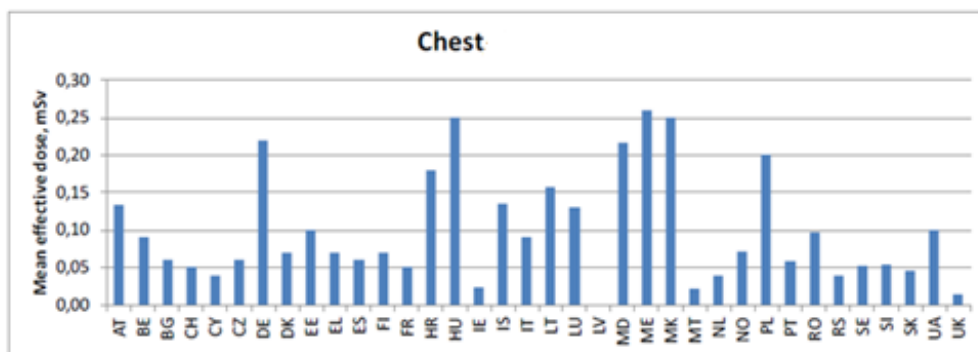


Figure 5-3: Typical effective doses (mSv) for chest radiography in different European countries [4]

The frequent use and diagnostic importance of chest X-ray makes it an interesting topic for optimization of image quality and patient dose. Therefore, the correlation between clinical and physical-technical image quality was assessed for digital chest radiography.

5.2 Paper I

Correlation of Contrast-Detail Analysis and Clinical Image Quality Assessment in Chest Radiography with a Human Cadaver Study

An De Crop¹ MSc, Klaus Bacher¹ PhD, Tom Van Hoof¹ PhD, Peter Smeets² MD, Barbara Smet² MD, Merel Vergauwen² MD, Urszula Kiendys² MD, Philippe Duyck² MD PhD, Koenraad Verstraete² MD PhD, Katharina D'Herde¹ MD PhD, Hubert Thierens¹ PhD

¹Ghent University, Department of Basic Medical Sciences, Proeftuinstraat 86, B-9000 Gent

²Ghent University Hospital, Department of Radiology, De Pintelaan 185, B-9000 Gent

Reprint from Radiology: 2012; 262 (1): 298-304

Advance in Knowledge: A statistically significant correlation ($r = 0.92$, $P < .001$) was observed between the visual grading analysis scores and the inverse image quality figures; results support the value of contrast-detail phantom analysis in the evaluation of clinical image quality in chest radiography.

Implication for Patient Care: The correlation between contrast-detail phantom analysis and clinical image quality assessment can be used to determine a contrast-detail reference curve, making it possible to optimize the quality of images obtained with digital chest radiography units.

ABSTRACT

Purpose: To determine the correlation between the clinical and physical image quality of chest images by using cadavers embalmed with the Thiel technique and a contrast-detail phantom.

Materials and Methods: The use of human cadavers fulfilled the requirements of the institutional ethics committee. Clinical image quality was assessed by using three human cadavers embalmed with the Thiel technique, which results in excellent preservation of the flexibility and plasticity of organs and tissues. As a result, lungs can be inflated during image acquisition to simulate the pulmonary anatomy seen on a chest radiograph. Both contrast-detail phantom images and chest images of the Thiel-embalmed bodies were acquired with an

amorphous silicon flat-panel detector. Tube voltage (70, 81, 90, 100, 113, 125 kVp), copper filtration (0.1, 0.2, 0.3 mm Cu), and exposure settings (200, 280, 400, 560, 800 speed class) were altered to simulate different quality levels. Four experienced radiologists assessed the image quality by using a visual grading analysis (VGA) technique based on European Quality Criteria for Chest Radiology. The phantom images were scored manually and automatically with use of dedicated software, both resulting in an inverse image quality figure (IQF). Spearman rank correlations between inverse IQFs and VGA scores were calculated.

Results: A statistically significant correlation ($r = 0.80$, $P < .01$) was observed between the VGA scores and the manually obtained inverse IQFs. Comparison of the VGA scores and the automated evaluated phantom images showed an even better correlation ($r = 0.92$, $P < .001$).

Conclusion: The results support the value of contrast-detail phantom analysis for evaluating clinical image quality in chest radiography.

INTRODUCTION

In digital radiography, patient radiation doses should be kept as low as possible while maintaining an appropriate image quality and accurate diagnostic value (1). Evaluation of image quality can be based on detector characteristics such as detective quantum efficiency, modulation transfer function, and contrast-to-noise ratio (2,3). These physical measurements describe the technical performance of the detector but are difficult to link directly to clinical performance (4). Contrast-detail phantom studies are well established in medical physics and can be used to compare the image quality of different radiography systems or different acquisition techniques (5–8). An advantage of the contrast-detail phantom technique is that it includes the complete imaging chain, including the human observer (9). However, because these phantom models are not related to patient anatomy, it is unclear whether this method is appropriate for simulating clinical image quality. Clinical image quality can be assessed by applying visual grading analysis (VGA) (10,11) or receiver operating characteristic (ROC) analysis (12–14) in a patient population. However, it is difficult to recruit patients for these studies because either large numbers of patient images must be available or one patient must be irradiated with different dose settings, which is ethically not allowed. As an alternative, clinical images of an anthropomorphic phantom can be acquired. Compared with contrast-detail phantoms, anthropomorphic phantoms enable a better approximation of the clinical reality with respect to anatomic background (9). In the present study, the image quality of

chest radiographs was assessed with use of human cadavers embalmed by using the Thiel technique (15). In contrast to the classic formol embalming technique, the Thiel embalming method results in excellent preservation of the flexibility and plasticity of organs and tissues (15,16). As a result, lungs can be inflated during image acquisition to simulate the anatomy on a chest radiograph. Because chest radiography involves many anatomic structures with different x-ray absorption, a wide range of gray values is observed on the clinical images. The dynamic range observed on radiographs of a contrast-detail phantom is typically much lower and, as a result, does not reflect the clinical reality. Therefore, the aim of the present study was to evaluate the correlation between the clinical and physical image quality of chest images by using cadavers embalmed with the Thiel technique and a contrast-detail phantom, respectively.

MATERIALS AND METHODS

Thiel embalmed Cadavers

The use of human cadavers fulfilled the requirements of our institutional ethics committee. Three human cadavers (one male, two female) were embalmed by using the Thiel method (16). In this technique, 4-chloro-3-methylenphenol, in addition to various salts, is used for fixation and boric acid is added for disinfection. Furthermore, ethylene glycol is used for preservation of tissue plasticity; the concentration of formalin is kept to the strict minimum (0.8%) (15). In contrast to the standard formalin embalming technique, this technique results in well-preserved organs and tissues with regard to color, consistency, natural flexibility, natural plasticity, and transparency. As a result, lung tissue is preserved completely, making it possible to ventilate the lungs by performing a tracheotomy in combination with balloon ventilation. After ventilating the lungs, chest radiographs can be acquired for subjective image quality analysis. An example of a chest radiograph obtained in a Thiel-embalmed cadaver is shown in Figure 1 .

Contrast-Detail phantom

The contrast-detail phantom (CDRAD 2.0; Artinis Medical Systems, Zetten, the Netherlands) consists of a Plexiglas plate (265 X 265 X 10 mm³) with a grid of 15 X 15 cells. The cells contain circular holes with depths and diameters varying logarithmically from 0.3 to 8.0 mm, simulating variations in contrast and resolution, respectively. For objects measuring 4 mm and smaller, two identical holes are present—one in the center and one placed at random in one

of the four cell corners. The presence of these additional objects enabled us to perform a four-alternative forced-choice experiment in which the observers must identify the locations of the corner objects if considered visible. In this way, these additional objects help minimize potential biases due to a priori knowledge of the presence of objects in every square region (7). This phantom was used to assess the minimum contrast required to visualize objects of different sizes above the noise threshold: the contrast-detail curve. The phantom was placed between two slabs of 5-cm-thick polymethylmethacrylate to simulate patient attenuation and scatter (17).



Figure 1: Example of a chest radiograph in a Thiel-embalmed cadaver

Image acquisition

For this study, the radiographic system (Siemens, Erlangen, Germany) consisted of an amorphous silicon flat-panel detector (Trixell, Moirans, France), x-ray tube (Optilix 150/30/50 HC-100, Siemens; focal spot size, 0.6 mm), and high-voltage generator (Polydoros LX 30 or 50 Lite, Siemens). The amorphous silicon image detector is equipped with a 43 X 43-cm² x-ray sensing surface with a 3000 X 3000 matrix and 143- μ m pixel size. Chest radiographs of the Thiel embalmed cadavers were acquired at a focus-to-detector distance of 150 cm by using automatic exposure control. Different settings of tube voltage (70, 81, 90, 100, 113, and 125 kVp), copper filtration (0.1, 0.2, and 0.3 mm Cu), and sensitivity (speed class of 200, 280, 400, 560, and 800) were applied. After lung ventilation, four radiographs of the Thiel-embalmed cadavers were obtained at each setting. Then, a series of radiographs of the contrast-detail phantom was acquired by using the same exposure settings as for the Thiel embalmed

cadavers. For the contrast-detail phantom experiment, 16 images were obtained per exposure setting.

Image analysis

After image acquisition, all data were sent to a picture archiving and communication system workstation (Centricity, version 2.0 CRS5 SP2; GE Healthcare, Barrington, Ill) for image quality assessment. Images were displayed on a 20-inch, 3-megapixel high-contrast grayscale monitor (model MFGD 3420; Barco, Kortrijk, Belgium). The monitor was calibrated to comply with the Digital Imaging and Communications in Medicine Part 3.14 Gray-Scale Standard Display Function, using calibration software provided by the manufacturer (MediCal Pro, Barco) (18). The readers were allowed to adjust image brightness and contrast and to magnify the images to full resolution. Images were presented in random order and interpreted independently.

Scoring of Cadaver Images

Four experienced radiologists (P.V.S., with 25 years of experience; K.V., with 24 years of experience; and B.S.S. and M.V., both with 6 years of experience) assessed the soft-copy chest radiographs and scored the image quality by using criteria adopted from the European Guidelines on Quality Criteria for Diagnostic Radiographic Images (19). The criteria are listed in Table 1. For all three cadavers, each image was viewed individually and each structure rated on a scale from 1 to 4 according to the criteria listed in Table 2. An absolute VGA score for each reader was calculated as follows :

$$VGAS = \frac{\sum_{s=1}^S \sum_{t=1}^O G_{abs,s,t}}{S * T}$$

where $G_{abs,s,t}$ is the rating for a particular structure (s) and Thiel-embalmed body (t) (20). S and T are the number of structures and cadavers, respectively. The latter scoring reflected the image quality of the individual images without the use of a reference image (21). Before starting the study, readers underwent a training session to familiarize themselves with the scoring method.

Table 1: Image quality criteria for chest radiography

Criterion number	Description
1	Medial border of the scapulae outside the lung fields
2	Reproduction of the whole rib cage above the diaphragm
3	Visually sharp reproduction of the vascular pattern in the whole lung, particularly the peripheral vessels
4	Visually sharp reproduction of the trachea and proximal bronchi
5	Visually sharp reproduction of the borders of the heart and aorta
6	Visually sharp reproduction of the diaphragm and lateral costo-phrenic angles
7	Visualization of the retrocardiac lung and the mediastinum
8	Visualization of the spine through the heart shadow
9	Small round details in the whole lung, including the retrocardiac areas: high contrast: 0.7 mm diameter
10	Small round details in the whole lung, including the retrocardiac areas: low contrast: 2 mm diameter
11	Linear and reticular details out to the lung periphery: high contrast: 0.3 mm in width
12	Linear and reticular details out to the lung periphery: low contrast: 2 mm in width

Table 2: Rating used to evaluate the clinical images

Rating	The structure in the image is:
1	Not visible
2	Poorly reproduced
3	Adequately reproduced
4	Very well reproduced

Scoring of Contrast-Detail Phantom Images

Four medical physicists identified the location of the corner holes in every square cell. The results were entered on a score sheet for each image reviewed. After the indicated positions of the corner holes were compared with the true hole positions in the phantom, a correction scheme was used, taking into account the nearest neighbors to get more accurate results. After the correction, the threshold contrast value ($C_{i,h}$) was determined for each different row (detail D_i) as the smallest visible depth in regions of valid detection (7). The inverse image

quality figure (IQF) was introduced for quantitative comparison of the contrast-detail images (22). The inverse IQF is defined as follows:

$$IQF_{inv} = \frac{100}{\sum_{i=1}^n C_i D_{i,th}}$$

where $D_{i,th}$ denotes the threshold diameter in contrast column i and C_i the correctly identified contrast values. The higher the inverse IQF, the better the low-contrast visibility. The inverse IQF was calculated for all analyzed images and averaged over the four readers. In addition to human reading, automated contrast-detail phantom analysis was performed with software (CDRAD Analyzer, version 2.1; Artinis Medical Systems). The CDRAD Analyzer analyzes the images and applies a statistical method to determine whether a certain contrast-detail combination is detected or not. This statistical method uses the average pixel signal value and standard deviation of both the image of the contrast-detail combination under evaluation and its background pixels. The program uses the Welch Satterthwaite method (Student t test with Welch correction) to determine if the average signal level in a certain square is greater than the average background level plus an “a priori difference of means” (23). If the difference between the two signals is statistically significant, at a specified significance level, the detail is detected. After evaluation of the image, a contrast-detail diagram is displayed and an inverse IQF calculated.

Statistical Analysis

Correlation between VGA scores and inverse IQFs was determined by means of the Spearman rank correlation coefficient with use of an ordinary leastsquare algorithm. Interobserver agreement for VGA scores and inverse IQFs was determined by calculating the intraclass correlation coefficient. An intraclass correlation coefficient of less than 0.8 was considered indicative of poor interobserver agreement. A 95% confidence interval was used for all statistical measures. All calculations were performed with software (PASW Statistics, version 18.0.3; SPSS, Chicago, Ill).

RESULTS

Figure 2 presents the automatically scored inverse IQFs versus the manually obtained inverse IQFs. The error bars in the x direction represent the standard deviation between the scores of the four independent readers. In the y direction, the error bars represent the standard deviation between the automatically obtained inverse IQFs of 16 images, taken at the same

exposure settings. A linear regression curve was plotted, and a Spearman rank correlation coefficient of 0.895 ($P < .001$) was obtained. The slope of the linear regression line showed a mean increase of 19% in the measurement of the automatically obtained inverse IQF compared with the value derived by the human observers. Therefore, the software tool was more sensitive than the human observers in the correct detection of objects in the contrast-detail phantom. There was good interobserver agreement, with intraclass correlation coefficients of 0.916 ($P < .001$) and 0.904 ($P < .001$) for VGA score and inverse IQF, respectively. As a result, mean scores were used for further analysis.

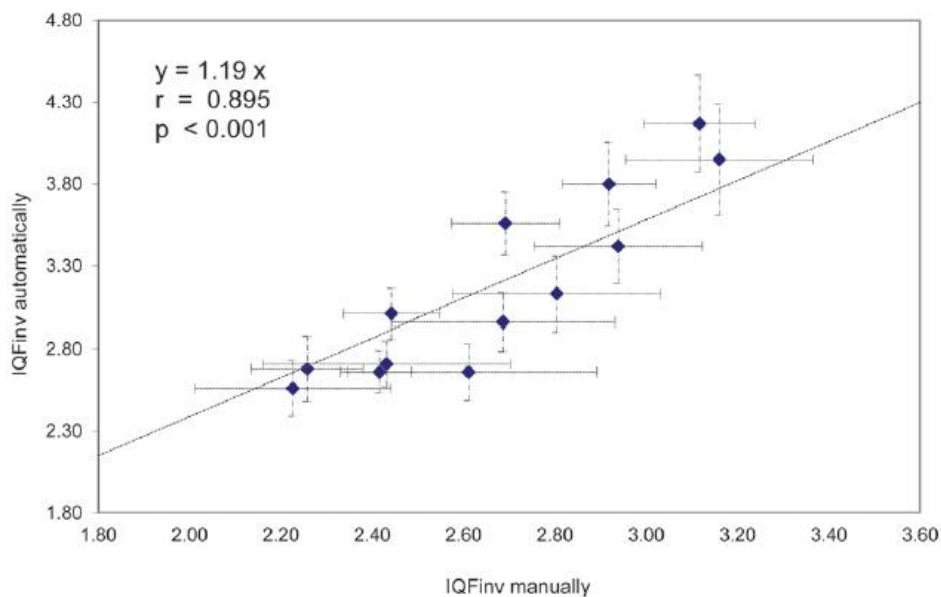


Figure 2: Graph shows automatically scored versus manually scored inverse IQF's. Error bars in x and y directions represent standard deviation between scores of four independent readers and between automatically obtained scores, respectively. Linear regression curve was plotted, and Spearman rank correlation coefficient was 0.895 ($P < .001$).

Table 3 gives an overview of all applied exposure parameters and the corresponding mean inverse IQF and VGA scores. For both inverse IQF and VGA score, an increase in image quality was noticed when lowering the kilovolt peak or sensitivity. Figure 3 shows the correlation between the mean VGA score of the four independent readers obtained with different exposure settings as a function of the automatically obtained inverse IQF from the contrast-detail images acquired with corresponding settings. The error bars represent the standard deviation between the scores of the different readers. A linear regression curve was plotted, and a Spearman correlation coefficient of 0.916 ($P < .001$) was found. For the manually

obtained inverse IQF, the correlation coefficient with the VGA score was 0.796 ($P = .001$). Further analysis showed that automatically obtained inverse IQFs had good correlation with different kilovolt peak and sensitivity settings. Spearman correlation coefficients of 0.999 ($P = .009$) and 0.900 ($P = .037$) were obtained for different kilovolt peak and sensitivity settings, respectively. For copper filtration, no significant correlation was found ($P = .051$). Analysis of the inverse IQF and VGA score data as a function of additional copper filtration revealed that the amount of copper filtration used (range, 0.1–0.3 mm Cu) did not have a significant influence on those parameters ($P = .443$ and $P = .509$, respectively).

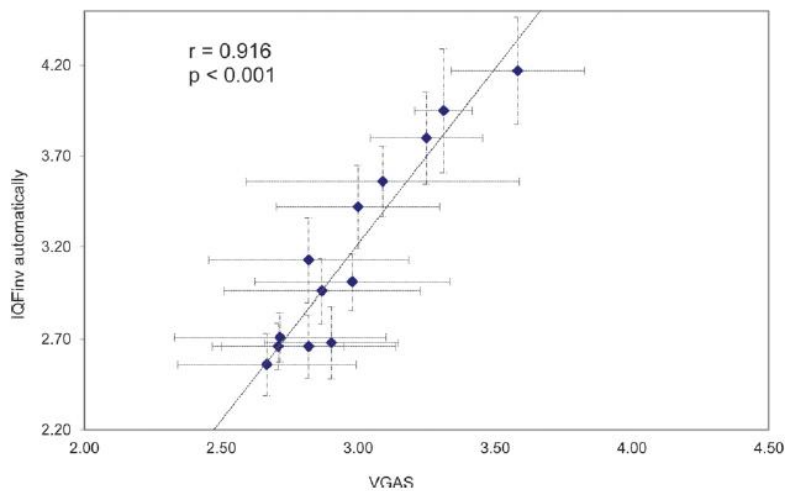


Figure 3: Graph shows mean VGA scores versus automatically obtained inverse IQF's. Error bars represent standard deviation between scores of different readers. Linear regression curve was plotted, and Spearman correlation coefficient was 0.916 ($P < .001$).

Table 3: Summary of the exposure parameters, IQFinv and VGAS

kVp	Cu filtration	Sensitivity	IQFinv \pm SD	VGAS \pm SD
125	0.1	400	2.68 \pm 0.20	2.90 \pm 0.25
125	0.2	400	2.66 \pm 0.17	2.82 \pm 0.32
125	0.3	400	2.66 \pm 0.13	2.71 \pm 0.24
70	0.0	400	4.17 \pm 0.30	3.58 \pm 0.24
81	0.0	400	3.95 \pm 0.34	3.31 \pm 0.11
90	0.0	400	3.80 \pm 0.26	3.25 \pm 0.21
100	0.0	400	3.56 \pm 0.20	3.09 \pm 0.50
113	0.0	400	2.96 \pm 0.18	2.84 \pm 0.36
125	0.0	400	3.01 \pm 0.15	2.98 \pm 0.36
125	0.0	200	3.42 \pm 0.23	3.00 \pm 0.30
125	0.0	280	3.13 \pm 0.23	2.82 \pm 0.37
125	0.0	560	2.71 \pm 0.14	2.72 \pm 0.39
125	0.0	800	2.56 \pm 0.17	2.67 \pm 0.33

DISCUSSION

The diagnostic information provided with digital detectors is equal or superior to that obtained with conventional screen-film systems (7). Because digital systems have a lot of practical and technical advantages, they are currently regarded as the gold standard in radiography. In contrast to screen-film systems, digital radiography units can still provide excellent image quality with patient overexposure (1). Therefore, appropriate optimization of digital radiography exposure is needed to avoid unnecessarily high patient doses (1). The latter optimization is an important task for the medical physics expert, as stated by the European Medical Exposure Directive (24). The main goal of optimizing digital radiography procedures is to determine the patient dose level required to provide sufficient image quality for making a correct diagnosis. Fortunately, because of their large dynamic range, digital radiography systems offer many possibilities for optimization. Methods of patient dose evaluation are easily available; however, techniques for image quality optimization are far more complicated. VGA and ROC studies are commonly used to assess clinical image quality. In VGA studies, relative or absolute scoring is performed on the basis of the visibility of normal anatomic structures (10,11,25). The task for observers in an ROC study is to detect whether a patient

contains a pathologic structure or not (12–14). However, these studies are difficult to implement in routine practice because they imply a substantial additional workload for the radiologists and large patient data groups must be available. Therefore, the latter methods are not feasible within a routine quality assurance program. A more practical approach to assessing image quality is the use of psychophysical measurements, where an observer has to recognize visual stimuli such as bar patterns or circular objects with different diameters and contrasts. These contrast-detail studies have been widely used for the objective analysis of the image quality performance of digital radiography systems (6,7,17,26). In contrast to VGA and ROC studies, contrast-detail studies are easily implemented in a quality assurance program because no patient data are required and images can be analyzed by the medical physics expert (27). However, the relationship between these contrast-detail studies and clinical image quality is not clearly defined. In fact, to be useful, these phantom measurements should be able to help predict changes in clinical image quality. The latter is not obvious because these phantoms do not simulate patient anatomy. As a result, the variations in gray values observed on images of contrast-detail phantoms are typically much lower than those of patients. In the current study, a VGA and contrast-detail study was set up to assess the relationship between contrast-detail analysis and clinical image quality assessment with use of Thiel-embalmed cadavers and a contrast-detail phantom. Cadavers fixed with conventional procedures, by using formalin for conservation, suffer from profound changes of color, strength, and fragility of organs and tissues. With use of the Thiel embalming method, the formalin content of the fixation solution is drastically reduced. With the low-odor embalming technique, the color, consistency, and transparency of organs and tissues are well preserved (15,16). Because this new technique also results in very good preservation of the lung structures, lungs can be inflated. To approximate as nearly as possible the normal patient anatomy, Thiel-embalmed bodies were ventilated during image acquisition. After the assessment of different thoracic regions by experienced radiologists, equivalency of thorax images of patients and Thiel embalmed cadavers was confirmed. This implies that instead of large patient groups, Thiel-embalmed bodies can be applied to assess clinical image quality by using VGA and ROC studies. Our study showed excellent correlation between contrast-detail parameters (inverse IQF) and clinically observed quality as scored by radiologists (VGA score). The correlation was even more pronounced when using automated computer analysis of the contrast-detail images. The latter can be attributed to the fact that there was no interobserver variability in the reading of the phantom images. The significant correlation between physical

and clinical image quality with use of a contrast-detail phantom and Thiel-embalmed cadavers confirms that differences in image quality seen in contrast-detail studies can actually help predict changes in the clinical image quality, thereby emphasizing the relevance of the contrast-detail method for optimization purposes. Only limited data describing the link between contrast-detail studies and clinical image quality are available in the literature. Geijer and Persliden (28) evaluated the image quality at different tube potential settings with use of anteroposterior lumbar spine radiography as a model. Image quality was assessed with VGA of an Alderson phantom and with CDRAD phantom contrast-detail analysis. They showed good consistency between results of the VGA and the computer reading of the contrast-detail analysis. However, the correlation between VGA and contrast-detail analysis was not further investigated. Bacher et al (17) performed a VGA and contrast-detail study with use of digital chest images to compare an amorphous silicon and amorphous selenium flat-panel detector. To obtain a contrast-detail performance equal to that of an amorphous silicon detector, they found it necessary to have a threefold increase in the entrance dose value used with the amorphous selenium detector system. The latter difference in entrance dose values was also observed in the clinical VGA study. However, no further correlation analysis was performed between VGA and the contrast-detail analysis. Image quality studies always have some limitations. In the current study, clinical image quality was assessed by means of a subjective overall quality score and not by means of detection of an abnormality. Detection of lesions by means of ROC analysis could give a more precise assessment of image quality for a specific clinical application. Tingberg et al (29) found that noise levels (and hence dose settings) had a significant effect on the VGA score, whereas the detection of abnormalities was not altered. As a result, doses could be further reduced on the basis of the pathology detection study, compared with the results of the VGA data. However, the required image quality may be considerably different depending on the anatomic region under investigation. Moreover, within a specific anatomic region, the quality will depend on the types of lesions to be detected (eg, chest tumors, interstitial nodular disease, interstitial linear disease). Further studies should reveal the influence of these factors. On the basis of our findings, further research can be performed to establish a reference contrast-detail curve for thorax radiography. The latter curve will simulate the minimum contrast-detail performance needed to obtain sufficient clinical quality in chest radiography. A similar reference curve already exists in the framework of digital mammography, as described in the European Reference Organization for Quality Assured Breast Screening and Diagnostic Services protocol for the

quality control of digital mammography systems (30). However, the latter curve was not determined on the basis of the link between clinical and physical findings. The availability of such reference curves is very useful within a quality assurance program to confirm if the local image quality fulfills the requirements. On the basis of the obtained results, settings can be changed to realize a sufficient level of image quality. The use of a reference contrast-detail curve is not limited to radiography of the thorax. Similar studies can be set up to determine a reference curve for other radiographic examinations (eg, abdominal or skeletal radiography).

Acknowledgements

The authors thank Aron De Smet, Bsc, for his contribution in the embalming procedure and the practical arrangements for the imaging of the cadavers.

References

1. International Commission on Radiological Protection . *Managing patient dose in digital radiology*. Publication 93 . Oxford, England : Elsevier , 2004 .
2. Båth M , Sund P , Månsson LG . *Evaluation of the imaging properties of two generations of a CCD-based system for digital chest radiography* . Med Phys 2002 ; 29 (10): 2286 – 2297 .
3. Samei E , Flynn MJ . *An experimental comparison of detector performance for direct and indirect digital radiography systems* . Med Phys 2003 ; 30 (4): 608 – 622 .
4. Månsson LG . *Methods for the evaluation of image quality: a review* . Radiat Prot Dosimetry 2000 ; 90 (1-2): 89 – 99 .
5. Veldkamp WJ , Kroft LJ , Boot MV , Mertens BJ , Geleijns J . *Contrast-detail evaluation and dose assessment of eight digital chest radiography systems in clinical practice* . Eur Radiol 2006 ; 16 (2): 333 – 341 .
6. Bacher K , Smeets P , Bonnarens K , De Hauwere A , Verstraete K , Thierens H . *Dose reduction in patients undergoing chest imaging: digital amorphous silicon flat-panel detector radiography versus conventional film-screen radiography and phosphor-based computed radiography* . AJR Am J Roentgenol 2003 ; 181 (4): 923 – 929 .
7. Rong XJ , Shaw CC , Liu X , Lemacks MR , Thompson SK . *Comparison of an amorphous silicon/cesium iodide flat-panel digital chest radiography system with screen/ film and computed radiography systems: a contrast-detail phantom study* . Med Phys 2001 ; 28 (11): 2328 – 2335 .
8. Geijer H , Beckman KW , Andersson T , Persliden J . *Image quality vs. radiation dose for a flat-panel amorphous silicon detector: a phantom study* . Eur Radiol 2001 ; 11 (9): 1704 – 1709 .
9. Veldkamp WJ , Kroft LJ , Geleijns J . *Dose and perceived image quality in chest radiography*. Eur J Radiol 2009 ; 72 (2): 209 – 217 .
10. Lanhede B , Båth M , Kheddache S , et al . *The influence of different technique factors on image quality of chest radiographs as evaluated by modified CEC image quality criteria* . Br J Radiol 2002 ; 75 (889): 38 – 49 .
11. Redlich U , Hoeschen C , Doehring W . *Assessment and optimisation of the image quality of chest-radiography systems* . Radiat Prot Dosimetry 2005 ; 114 (1-3): 264 – 268 .

12. Leppert AG , Prokop M , Schaefer-Prokop CM , Galanski M . *Detection of simulated chest lesions: comparison of a conventional screenfilm combination, an asymmetric screen-film system, and storage phosphor radiography* . Radiology 1995 ; 195 (1): 259 – 263 .
13. Samei E , Flynn MJ , Peterson E , Eyler WR . *Subtle lung nodules: influence of local anatomic variations on detection* . Radiology 2003 ; 228 (1): 76 – 84 .
14. van Heesewijk HP , van der Graaf Y , de Valois JC , Feldberg MA . *Effects of dose reduction on digital chest imaging using a selenium detector: a study of detecting simulated diffuse interstitial pulmonary disease* . AJR Am J Roentgenol 1996 ; 167 (2): 403 – 408 .
15. Thiel W . *The preservation of the whole corpse with natural color* . Ann Anat 1992 ; 174 (3): 185 – 195 .
16. Thiel W . *Supplement to the conservation of an entire cadaver according to W. Thiel* . Ann Anat 2002 ; 184 (3): 267 – 269 .
17. Bacher K , Smeets P , Vereecken L , et al . *Image quality and radiation dose on digital chest imaging: comparison of amorphous silicon and amorphous selenium flat-panel systems* . AJR Am J Roentgenol 2006 ; 187 (3) : 630 – 637 .
18. American Association of Physicists in Medicine . *Specification and acceptance testing of computed tomography scanners. Report no. 39. College Park, Md : American Association of Physicists in Medicine , 1993 .*
19. European Commission . *European guidelines on quality criteria for diagnostic radiographic images (EUR 16260). Luxembourg : Office for Official Publications of the European Communities , 1996 .*
20. European Commission . *CEC quality criteria for diagnostic radiographic images and patient exposure trial (EUR 12952 EN). Luxembourg : Office for Official Publications of the European Communities , 1990 .*
21. Sund P , Båth M , Kheddache S , Månsson LG . *Comparison of visual grading analysis and determination of detective quantum efficiency for evaluating system performance in digital chest radiography* . Eur Radiol 2004 ; 14 (1): 48 – 58 .
22. Thijssen M , Bijkerk K , van der Burgh R . *Manual contrast-detail phantom CDRAD type 2.0. project quality assurance in radiology. St Radboud, the Netherlands : Department of Radiology, University Hospital Nijmegen , 1998 .*
23. Artinis Medical Systems . *Manual CDRAD 2.0 phantom and analyser software version 2.1. Zetten, the Netherlands : Artinis Medical Systems , 2009 .* 24 . *The Council of the European Union . Council directive 97/43/Euratom . Official Journal of the European Communities 1997 ; L (180): 22 – 27 .*
25. van Heesewijk HP , Neitzel U , van der Graaf Y , de Valois JC , Feldberg MA . *Digital chest imaging with a selenium detector: comparison with conventional radiography for visualization of specific anatomic regions of the chest* . AJR Am J Roentgenol 1995 ; 165 (3): 535 – 540 .
26. De Hauwere A , Bacher K , Smeets P , Verstraete K , Thierens H . *Analysis of image quality in digital chest imaging* . Radiat Prot Dosimetry 2005 ; 117 (1-3): 174 – 177 .
27. Cook LT , Insana MF , McFadden MA , Hall TJ , Cox GG . *Comparison of the lowcontrast detectability of a screen-film system and third generation computed radiography* . Med Phys 1994 ; 21 (5): 691 – 695 .
28. Geijer H , Persliden J . *Varied tube potential with constant effective dose at lumbar spine radiography using a flat-panel digital detector* . Radiat Prot Dosimetry 2005 ; 114 (1-3): 240 – 245 .

-
29. Tingberg A , Båth M , Håkansson M , et al . *Evaluation of image quality of lumbar spine images: a comparison between FFE and VGA* . Radiat Prot Dosimetry 2005 ; 114 (1-3): 53 – 61 .
 30. van Engen R , Young K , Bosmans H , Thijssen M . *European guidelines for quality assurance in breast cancer screening and diagnosis* . 4th ed. Luxembourg : Office for Official Publications of the European Communities , 2006 .

Correlation of Clinical and Physical-Technical Image Quality in Chest CT: A Human Cadaver Study Applied on Iterative Reconstruction

6.1 Background

6.1.1 CT

Computed tomography (CT) was first introduced in the early 1970s by G.N. Hounsfield. The main application of CT is transversal slices or 3D imaging instead of 2D projection imaging. It has important advantages compared to conventional radiography: X-ray attenuation differences of less than 1% can be detected and superimposition of structures is eliminated. While in conventional imaging, the attenuation difference needs to be more than 2% to become visible and organs are superimposed to on other.

CT is an imaging technique based on the reconstruction of the linear X-ray attenuation coefficient as a function of spatial coordinates in the transversal imaging plane. To perform this task, the attenuation of an X-ray beam crossing the object or patient has to be measured from a large number of different angles [6]. Afterwards, a reconstruction filter or reconstruction kernel is applied. These are typically divided in bone or sharp kernels and soft tissue or smooth kernels. Sharp kernels accentuate higher frequencies in the image at the

expense of increased noise while smooth kernels produce images with reduced noise but lower spatial resolution [7]. CT scanning results usually in transversal slices, although sagittal and coronal sections can also be constructed by digital processing of the transversal images [8]. Combinations of the different orientation leads to 3D imaging, used in radiotherapy planning.

Since their development in the early 70s, CT devices have constantly been improved. Modern CT scanners facilitate faster gantry rotation cycle times, helical scanning technique, thinner detector widths, increased numbers of detector rows, improved image quality due to increased spatial resolution and decreased motion artifacts and finally faster data acquisition with wider volume coverage [9]. All these advantages have greatly extended the role of CT imaging in medical X-ray diagnostics.

The radiation dose in CT examinations is typically expressed in the Computed Tomography Dose Index (CTDI). The CTDI represents the average absorbed dose, along the z-axis from one axial CT scan. The CTDI is calculated by dividing the integrated absorbed dose by the nominal total beam collimation [10]. To represent dose for a specific scan protocol, which almost always involves a series of scans, it is essential to take into account any gaps or overlaps between the x-ray beams from consecutive rotations of the x-ray source as expressed in the pitch factor. This is accomplished with use of a dose descriptor known as the Volume CTDI (CTDI_{vol}). The CTDI_{vol} provides a single CT dose parameter, based on a directly and easily measured quantity, which represents the average dose within the scan volume for a standardized CTDI phantom[10].

6.1.2 Iterative reconstruction

CT images are reconstructed from projection data, which represent the total integrated attenuation of the X-ray. The image reconstruction technique uses these projection data to generate a CT image such that each pixel value ideally represents the attenuation of the patient at that pixel location [8]. Since the introduction of CT, filtered back projection (FBP) has been the primary image reconstruction technique. The standard FBP algorithm is a compromise between reconstruction speed and image noise. Fast reconstruction times comes at the expense of ignoring scanner hardware-specific and photon noise statistic information, which introduces higher image noise and artifacts in CT images [11]. Although FBP is a standard image reconstruction method in CT, there is a trade-off between spatial resolution

and image noise. Hence, the amount of radiation dose reduction is limited by the diagnostic quality of images [12].

With the goal of reducing patient dose and at the same time maintaining the diagnostic image quality, iterative reconstruction (IR) algorithms were implemented in the last generation of CT equipment. Unlike analytical reconstruction algorithms such as FBP, IR generates a set of synthesized projections by accurately simulating and modeling the data collection process in CT. The model incorporates a physical model of the CT system that includes scanner geometric information (including size of each detector cell, dimensions of the focal spot and the shape and size of each image voxel) and system statistical information (including photon statistics and electronic noise in the data acquisition system) [13]. While FBP is based on only a single reconstruction, IR algorithms use multiple repetitions. As a consequence, the computational demands are much higher. Due to the exponential growth of computer power, the use of IR methods has become a realistic option with reconstruction times acceptable for clinical workflow.

Essentially, iterative reconstruction introduces a correction loop in the image generation process that cleans up artifacts and noise. After the initial reconstruction, artificial raw data are calculated from the reconstructed image, as if the image had been the measurement object in a CT scan. These new data are compared to the actually measured raw data, and a correction image is calculated (*Figure 6-1*). The iteration process can be initiated using prior information, for example, a standard FBP reconstruction [14]. In general, the better the prior images match the final images, the faster the process leads towards a stable solution. The iterative process is finished when either a fixed number of iterations is reached, or the update for the current image estimate is considered close enough to the previous one or when a predefined quality criterion in the image estimate is fulfilled.

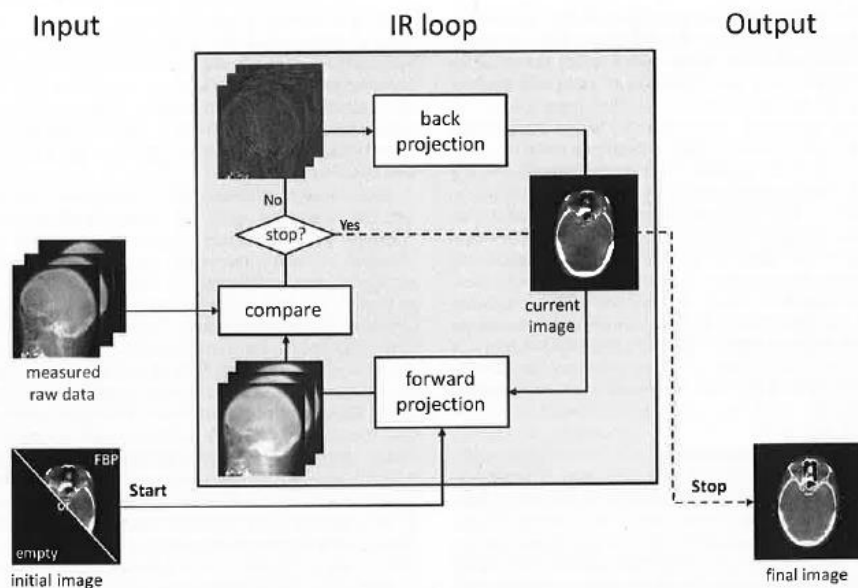


Figure 6-1: Schematic view of the iterative reconstruction process [14]

Most IR techniques work in the raw data domain. However, it is also possible to incorporate a correction loop in the image domain [11]. The only real advantage of the raw data loop is correction for artifacts caused by the inexact nature of FBP reconstructions. However, the forward projection is the most complex and the most computationally expensive step of the IR process. Therefore, it makes sense to use only the loop in the raw data domain when artifact correction is necessary [15]. Otherwise, the use of the raw data loop would only result in increased reconstruction time but not improve image quality compared to the use of the image based correction loop.

In 2008 General Electric (Milwaukee, WI, USA) introduced the first CT iterative reconstruction method commercially available for clinical applications, under the name of adaptive statistical iterative reconstruction (ASIR). ASIR uses a blend of filtered back projection images with iteratively reconstructed images in the raw data domain to reduce the image noise [16]. Two years later, Siemens (Erlangen, Germany) released an iterative reconstruction in image space (IRIS). Contrary to ASIR, IRIS is based on an iterative reconstruction loop in the image domain to speed up the reconstruction process [17]. Sinogram affirmed iterative reconstruction (SAFIRE, Siemens Healthcare) takes information from raw data and processes in the image domain to compensate for longer reconstruction time [15]. In 2011, Philips commercialized the fourth version of its iterative method which was called iDose. After noise removal in the raw data domain, an optimal anatomical model is used in the image domain to iteratively

eliminate the quantum image noise and to maintain the appearance of a full dose image [18]. Toshiba has also developed an iterative method, adaptive iterative dose reduction (AIDR), which works both in the image and in the raw data domain and automatically calculates the optimized number of iterations [19]. GE introduced a full iterative reconstruction or model-based iterative reconstruction (MBIR- under the commercial name of Veo). Unlike ASIR, Veo is a fully iterative method working in the raw data domain, which takes not only the data statistics into account but also the geometry of the CT equipment itself [20].

6.1.3 Thorax CT

Compared with chest radiography, CT has greater diagnostic accuracy in a range of clinical situations. In 2011, the frequency of chest CT examinations in Belgium was 30 per 1000 of population [4]. In the same period, the average amount of chest CT examinations over 36 European countries was 13 per 1000 of population [4] (*Figure 6-2*). In 2012 and 2013, about 260 000 chest CT examinations, i.e. 12% of all CT examinations, were performed in Belgium [5].

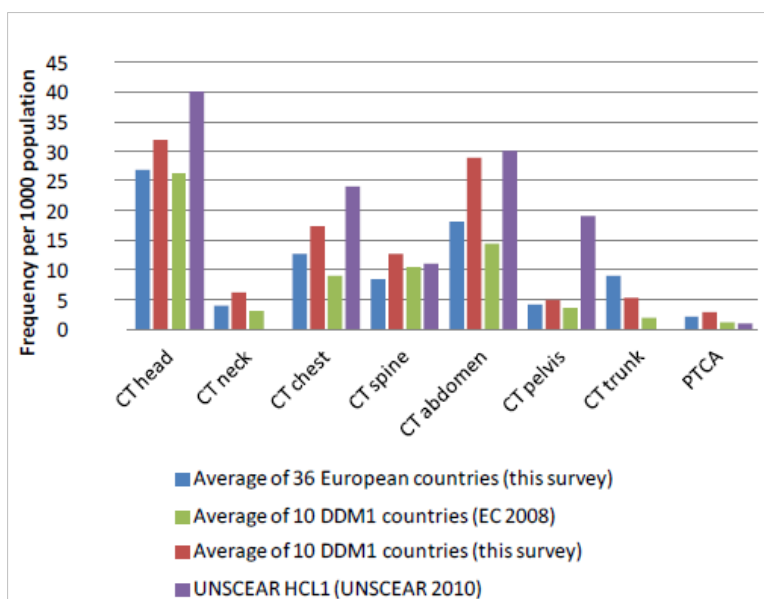


Figure 6-2: Average frequencies per 1000 of population for different CT examinations [4]

The mean effective dose per patient estimated in European countries for chest CT examinations is 6.6 mSv, in Belgium this is 4.2 mSv [4] (*Figure 6-3*).

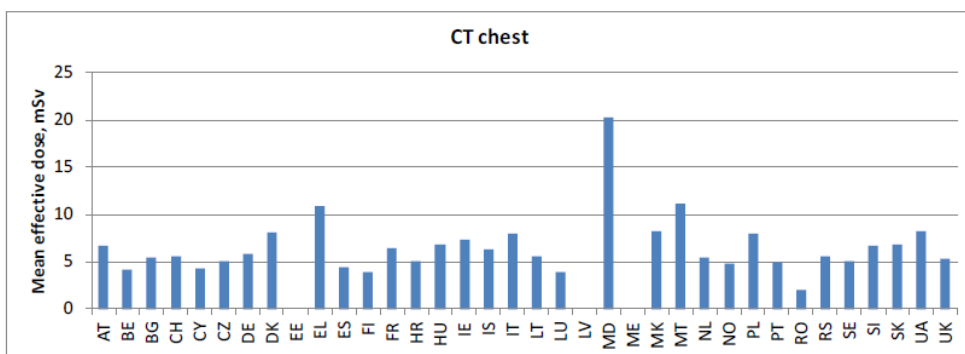


Figure 6-3: Typical effective doses (mSv) for chest CT in different European countries [4]

The mean effective dose associated with CT is remarkably higher than the mean effective dose associated with plain radiography. Especially in chest imaging, where the effective dose of CT is 50 times higher than for a chest radiography. Therefore, the main challenge of CT is to keep radiation dose to a minimum, while still obtaining diagnostic image quality. For this reason, several new image acquisition and reconstruction techniques have been developed aiming to reduce radiation exposure.

6.2 Paper II

Correlation of clinical and physical-technical image quality in chest CT: a human cadaver study applied on iterative reconstruction.

An De Crop¹, Peter Smeets², Tom Van Hoof¹, Merel Vergauwen², Tom Dewaele², Mathias Van Borsel², Eric Achten², Koenraad Verstraete², Katharina D'Herde¹, Hubert Thierens¹, Klaus Bacher¹

¹Ghent University, Department of Basic Medical Sciences, Proeftuinstraat 86, B-9000 Gent

²Ghent University Hospital, Department of Radiology, De Pintelaan 185, B-9000 Gent

Reprint from BMC Medical Imaging: 2015; 15 (1): 32

ABSTRACT

Background: The first aim of this study was to evaluate the correlation between clinical and physical-technical image quality applied to different strengths of iterative reconstruction in chest CT images using Thiel cadaver acquisitions and Catphan images. The second aim was to determine the potential dose reduction of iterative reconstruction compared to conventional filtered back projection based on different clinical and physical-technical image quality parameters.

Methods: Clinical image quality was assessed using three Thiel embalmed human cadavers. A Catphan phantom was used to assess physical-technical image quality parameters such as noise, contrast-detail and contrast-to-noise ratio (CNR).

Both Catphan and chest Thiel CT images were acquired on a multislice CT scanner at 120kVp and 0.9 pitch. Six different refmAs settings were applied (12, 30, 60, 90, 120 and 150refmAs) and each scan was reconstructed using filtered back projection (FBP) and iterative reconstruction (SAFIRE) algorithms (1,3 and 5 strengths) using a sharp kernel, resulting in 24 image series. Four radiologists assessed the clinical image quality, using a visual grading analysis (VGA) technique based on the European Quality Criteria for Chest CT.

Results: Correlation coefficients between clinical and physical-technical image quality varied from 0.88 to 0.92, depending on the selected physical-technical parameter. Depending on the

strength of SAFIRE, the potential dose reduction based on noise and CNR and the inverse image quality figure (IQF_{inv}) varied from 14.0% to 67.8%, 16.0 to 71.5% and 22.7% to 50.6% respectively. Potential dose reduction based on clinical image quality varied from 27% to 37.4%, depending on the strength of SAFIRE.

Conclusion: Our results demonstrate that noise assessments in a uniform phantom overestimate the potential dose reduction for the SAFIRE IR algorithm. Since the IQF_{inv} based dose reduction is quite consistent with the clinical based dose reduction, an optimized contrast-detail phantom could improve the use of contrast-detail analysis for image quality assessment in chest CT imaging. In conclusion, one should be cautious to evaluate the performance of CT equipment taking into account only physical-technical parameters as noise and CNR, as this might give an incomplete representation of the actual clinical image quality performance.

BACKGROUND

The number of CT examinations has increased rapidly over the last few years, resulting in a substantial increase in radiation dose of the population in the Western world [1]. It has been estimated that these CT examinations may be responsible for approximately 2% of all incident cancer cases in the United States [2]. Consequently, a lot of efforts have been made over the last decade to reduce the radiation dose for the patient by introducing new techniques such as automatic tube current modulation, adaptive collimation and iterative reconstruction [3-6]. If new dose reduction techniques are implemented, the impact on the image quality has to be investigated.

Medical physicists assess the image quality in CT using technical phantoms, evaluating parameters as noise, modulation transfer function (MTF), contrast-to-noise ratio (CNR) and/or contrast-detail. However, as these phantom models are not related to patient anatomy, it is unclear whether this methodology is appropriate to evaluate the clinical image quality. Particularly for noise, this can be problematic, since noise measurements in a uniform phantom don't account for the complex relationship between anatomical variability and image quality [7]. To be able to compare the performance of different CT scanners or to evaluate dose optimization tools, it is of critical importance that physical-technical image quality based dose optimization performance is related to the clinical image quality based dose optimization performance.

Clinical image quality is typically assessed by applying a visual grading analysis (VGA) [8] or a receiver operating characteristics (ROC) [9] study setup in a patient population. However, these patient studies are rather difficult to implement since either large numbers of patient images must be available or one patient has to be exposed to different dose settings, which should be avoided from ethical point of view. As an alternative, clinical images of an anthropomorphic phantom can be acquired. Compared to physical-technical phantoms, these phantoms approximate better the clinical reality with respect to anatomical features [10].

In present study, patient image quality of chest CT was assessed by means of human cadavers, conserved using the Thiel embalming technique [11]. In contrast to the classical formol embalming technique, the Thiel embalming method results in excellent preservation of the flexibility and plasticity of organs and tissues [11, 12]. As a result, lungs can be inflated during image acquisition to simulate the anatomy of a chest CT [13]. Consequently, these Thiel embalmed cadavers are an excellent model to investigate the link between clinical and physical-technical image quality. This link was already established in conventional chest radiography [13]. However, with respect to CT imaging, the correlation between clinical and physical-technical image quality was not yet examined.

The first aim of this study was to evaluate the correlation between clinical and physical-technical image quality applied to different strengths of iterative reconstruction in chest CT images using Thiel cadaver acquisitions and Catphan images. The second aim was to determine the potential dose reduction of iterative reconstruction compared to conventional filtered back projection based on different clinical and physical-technical image quality parameters.

MATERIALS AND METHODS

Thiel embalmed cadavers

The use of human cadavers is in compliance with the Helsinki Declaration and fulfilled the requirements of the ethical committee of our institution (Ghent University, B67020095736). The cadavers were obtained from body donations to the department of Anatomy of Ghent University. Three human cadavers (2 male, 1 female) were embalmed using the methodology of Prof. Em. Walther Thiel, Anatomisches Institut Karl-Franzens-Universität, Graz, Austria [12]. Hereby, 4-chloro-3-methylenphenol as well as various salts are used for fixation and boric acid is added for disinfection. Furthermore, ethylene glycol is used for preservation of tissue plasticity, while the concentration of formalin is kept to the strict minimum (0.8%) [11]. In

contrast to standard formalin-embalmed human cadavers, this technique results in well preserved organs and tissues concerning color, consistency, natural flexibility and natural plasticity. As a result, lung tissue is preserved completely which makes it possible to ventilate the lungs by performing a tracheotomy in combination with balloon ventilation. After ventilating the lungs, chest CT acquisitions can be acquired for subjective image quality analysis. In the cadavers used in this study, the lungs showed signs of pulmonary oedema and pulmonary parenchymal consolidation. Equivalency of patient and Thiel thoracic CT images is displayed in Figure 1.

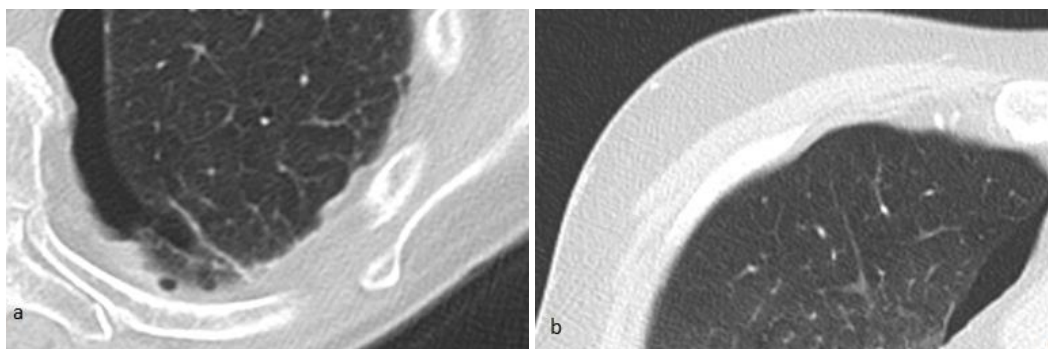


Figure1 : Patient versus Thiel cadaver chest CT image. Normal lung parenchyma illustrating nodular hypodense structures in a low density area, nodular hyperdense structures in a low density area, inter- or intralobular septa and the visceral pleura in both a patient (a) and a Thiel cadaver (b) chest CT image.

Catphan phantom

To evaluate the physical-technical image quality the Catphan@504 phantom (The Phantom laboratory, Salem, New York, USA) was used. The phantom consists of several modules to evaluate high and low contrast resolution, CNR and noise (Figure 2). In the low contrast module there are three areas with different contrast levels: 1%, 0.5% and 0.3%. Each contrast level contains targets with decreasing diameters (15, 9, 8, 7, 6, 5, 4, 3 and 2 mm). The CT number linearity and CT number accuracy module contains targets made from teflon, delrin, acrylic, polystyrene, low density polyethylene (LDPE), polymethylpentene (PMP) and air. The image uniformity module is made from a uniform material. The material's CT number is designed to be within 2% (20HU) of water's density at standard scanning protocols.

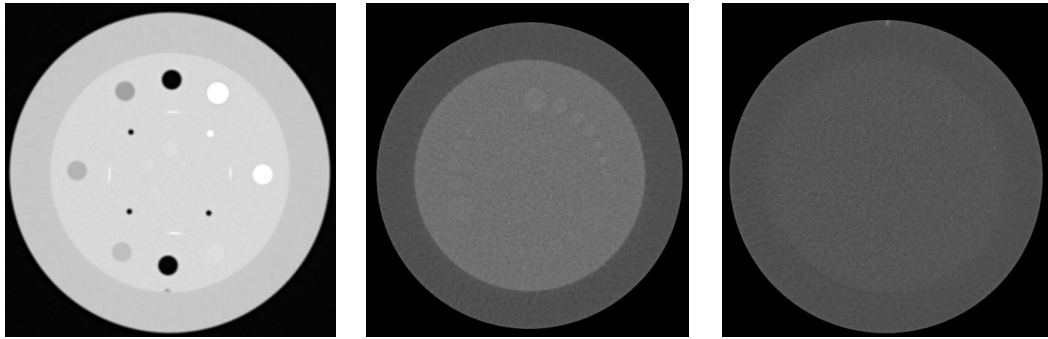


Figure 2: Catphan@504 phantom. The figure represents a CT image of the Catphan phantom. On the left, the CT number linearity and CT number accuracy module, which includes samples of teflon and acrylic used to calculate the CNR. In the middle, the low contrast module containing targets with different contrast levels: 1%, 0.5% and 0.3%. Each contrast level has 9 targets with different diameters: 15, 9, 8, 7, 6, 5, 4, 3, 2 mm. On the right the image uniformity module used to evaluate the noise using a circular ROI.

Image acquisition

All images in this study were acquired with a Somatom Definition Flash CT scanner (Siemens Healthcare, Erlangen, Germany). The CT scanner is equipped with the dual source technology, CARE Dose4D, CARE kV, and Sinogram Affirmed Iterative Reconstruction (SAFIRE).

Chest CT scans of the lung ventilated Thiel embalmed cadavers were acquired using CARE Dose4D at different reference mAs values (12, 30, 60, 90, 120 and 150 refmAs), resulting in a mean $CTDI_{vol}$ of 0.84, 2.05, 4.08, 6.18, 8.35 and 11.59 mGy respectively. The 90 refmAs setting is clinically applied in our institution. Other scan parameters were 120kVp and pitch 0.9. Each data set was reconstructed at 3 mm using filtered back projection (FBP) with a sharp kernel (B70). To compare the FBP and the SAFIRE technique, all six data sets were also reconstructed using different strengths of IR (1,3 and 5 iteration steps). Similarly to the FBP reconstructed images, IR images were reconstructed using a sharp kernel (I70-1, I70-3, I70-5), resulting in a total of 24 image series (6 refmAs settings with each 4 reconstruction settings).

Afterwards Catphan images were acquired without CARE Dose4D at a mAs value corresponding to the mean mAs value over the different slices in the Thiel cadaver acquired at the six different refmAs settings, resulting in a $CTDI_{vol}$ of 0.84, 2.11, 4.19, 6.37, 8.82 and 12.23 mGy respectively. The same reconstruction settings as for the Thiel embalmed cadavers were used. All scanning and reconstruction parameters and the investigated phantoms and image quality parameters are listed in Table 1.

Table 1: Scanning and reconstruction parameters, investigated phantoms and image quality parameters used in this study

Fixed scan parameters	CTDI _(vol) Thiel / Catphan	Reconstruction parameters for each CTDI _(vol)	Scanned objects	Investigated image quality parameters	
				Thiel	Catphan
120kVp	0.84 / 0.84	B70	Thiel cadavers (3)	VGAS	Noise
0.9 pitch	2.11 / 2.05	I70/1	Catphan phantom		IQF _{inv}
3 mm reconstruction thickness	4.19 / 4.08	I70/3			CNR
	6.37 / 6.18	I70/5			
	8.82 / 8.35				
	12.23 / 11.59				

Image quality analysis

After acquisition, all data were sent to a PACS Workstation (GE Centricity PACS version 2.0 CRS5 SP2) for image quality assessment. Images were displayed on a 30-inch, 3-megapixel high-contrast color monitor (Barco MDCC 6130DL, Kortrijk, Belgium). The monitor was calibrated to comply with the DICOM Part 3.14 Greyscale Standard Display Function, using calibration software provided by the manufacturer (MediCal Pro, BARCO, Kortrijk, Belgium) [14]. Maximum luminance of all monitors was adjusted to 400cd/m² and ambient lighting levels were below 50 lux as recommended by AAPM TG 18 [15].

Scoring of Thiel images

Four experienced radiologists (PS: 25 years of experience; TDW, MVB and MV: 6 years of experience) assessed the chest CT scans and scored the image quality using criteria based on the European Guidelines on Quality Criteria for Computed Tomography [16]. The criteria are listed in Table 2.

Table 2: Image quality criteria for chest CT

Criterion No.	Description:
1	Visually sharp reproduction of a nodular hypodense structure in a high density area such as an alveolus in consolidated lung parenchyma
2	Visually sharp reproduction of a nodular hypodense structure in a low density area such as normal lung parenchyma
3	Visually sharp reproduction of a nodular hyperdense structure in a low density area such as a vessel in aerated lung parenchyma
4	Visually sharp reproduction of an inter- or intralobular septum
5	Visually sharp reproduction of the bronchial wall
6	Visually sharp reproduction of the lung fissure
7	Visually sharp reproduction of a peripheral pulmonary artery branch
8	Visually sharp reproduction of fibrous strands
9	Visually sharp reproduction of the parietal and or visceral pleura

All criteria were evaluated in a predefined image area and a predefined image slice. For all three Thiel bodies, each stack was viewed individually and each structure was rated on a scale from 1 to 4 according to Table 3. An absolute VGA score (VGAS) for each reader was calculated as:

$$VGAS = \frac{\sum_{s=1}^S \sum_{t=1}^T G_{abs,s,t}}{S * T} \quad [17]$$

where $G_{abs,s,t}$ is the rating for a particular structure (s) and Thiel body (t). S and T are the number of structures and Thiel body's, respectively 9 and 3. The latter scoring reflected the image quality of the individual images without using a reference image [18].

All series were evaluated by the radiologists using Viewdex [19], a Java-based DICOM-compatible software tool for presentation and evaluation of images, without influencing the image quality. All images were blinded for acquisition and reconstruction parameters. The readers were allowed to adjust the image brightness and contrast and to magnify the images to full resolution. Viewdex defines a random order for each individual reader and all stacks were interpreted independently.

Before starting the study, a training session was organized to familiarize the readers with the scoring methodology.

Table 3: Rating used to evaluate the clinical images

Rating	The structure in the image is:
1	Not visible
2	Poorly reproduced
3	Adequately reproduced
4	Very well reproduced

Scoring of the Catphan phantom

Six medical physicists identified the minimally visible target diameter at three different contrast levels. The inverse image quality figure (IQF_{inv}) was introduced for quantitative comparison of the contrast-detail images [20]. The inverse image quality figure is defined as

$$IQF_{inv} = \frac{100}{\sum_{i=1}^n C_i D_{i,th}}$$

where $D_{i,th}$ denotes the threshold diameter for contrast i in mm, and C_i denotes the contrast value. The higher the IQF_{inv} , the better the low-contrast visibility. The IQF_{inv} was calculated for all analyzed images and averaged over the six readers.

The contrast to noise ratio relative to acrylic (soft tissue equivalent material) for teflon (bone equivalent material) was defined as:

$$CNR = \frac{(ROI_t - ROI_a)}{SD_a}$$

where ROI_t is the mean attenuation for teflon, ROI_a the mean attenuation for acrylic and SD_a the mean noise for acrylic. CT attenuation values and mean noise (in Hounsfield units) for teflon and acrylic were obtained by manually placing a circular region of interest (ROI) of 200 pixels in the target materials. CNR's were calculated in four consecutive slices of the Catphan CT number linearity and CT number accuracy module.

The image noise was evaluated using a circular ROI of 230 x 230 pixels in 11 following slices in the Catphan uniformity module.

Statistical analysis

To determine the influence of different exposure and reconstruction settings, data were analyzed using the Friedman test, a signed rank, non-parametric test used when comparing more than two related samples.

Inter-observer agreement for VGAS and IQF_{inv} values was determined by calculating the intraclass correlation coefficient. An intraclass correlation coefficient greater than 0.9 was considered to suggest an excellent inter-observer agreement [21].

After analysis of different fitting curves, a power function was selected as the best possible fit. Power functions are plotted for VGAS, noise, IQF_{inv} and CNR as a function of the mAs value. These curves are used to calculate the potential dose reduction when changing from a filtered back projection kernel (B70) at the clinically applied 90 refmAs to an iterative reconstruction kernel while maintaining the same value for noise, contrast-detail or CNR. To obtain a significant dose reduction, the two curves that are used, should differ significantly. This was examined by means of a Wilcoxon test, a signed rank, non-parametric test used when comparing two related samples. For this, all different readings (4, 11, 6, and 4 for VGAS, noise, IQF_{inv} and CNR) for the six different mAs settings are considered which result in 24, 66, 36 and 24 data points for VGAS, noise, IQF_{inv} and CNR respectively.

A 95% confidence interval was used for all statistical measures. All calculations were performed using the SPSS software tool (IBM SPSS statistics 22, IBM corp., NY, USA).

RESULTS

Excellent inter-observer agreement among the participating radiologists and among medical physicists was found by means of an intraclass correlation coefficient of 0.919 ($p < 0.001$) and 0.951 ($p < 0.001$) for the VGAS and IQF_{inv} parameters respectively. As a result, in the further analysis, scores averaged over the readers were used.

To evaluate the correlation between clinical and physical-technical image quality, regression curves were plotted for noise, CNR and IQF_{inv} as a function of VGA scores for the different refmAs settings (Fig 3). Good correlation was found between noise and VGAS, 0.90, $p < 0.001$. A correlation coefficient of 0.88, $p < 0.001$ was obtained for CNR and VGAS. Contrast-detail (IQF_{inv}) and VGAS resulted in a correlation coefficient of 0.92, $p < 0.001$.

To examine the influence of the iterative reconstruction strengths, the reconstruction settings mentioned in the materials and methods were applied to the Thiel images at 90 ref mAs.

Catphan images acquired at mAs settings corresponding to 90 ref mAs were selected. A significant effect of the IR strengths was found for both the physical-technical and clinical image quality parameters ($p < 0.05$) except for IQF_{inv} ($p = 0.706$).

For both clinical and physical-technical image quality parameters as a function of the mAs value, a power function fit was applied for all types of kernels (noise and CNR $r^2 > 0.9$, VGAS and IQF_{inv} $r^2 > 0.8$, $p < 0.05$). As expected, for all 4 different types of reconstruction kernel, a significant effect of mAs settings was confirmed by means of a Friedman test for noise and contrast detail ($p < 0.001$) and for CNR ($p < 0.05$). Correspondingly, this influence was also found for VGAS ($p < 0.05$).

A significant difference was found between the curve of the B70 kernel and the curve of each strength of iterative reconstruction for all clinical and physical-technical image quality parameters, except for VGAS B70-I70/1.

The power function for VGAS, noise, CNR and IQF_{inv} as a function of refmAs settings is shown in Figure 4, 5, 6 and 7. These curves were used to calculate the potential dose reduction when changing from a filtered back projection kernel at the clinically applied 90 refmAs to an iterative reconstruction kernel while maintaining the same value for noise, CNR or contrast-detail. In general, higher strengths of SAFIRE result in higher potential dose reduction. The potential dose reduction based on noise and CNR and IQF_{inv} varied from 14.0% to 37.8%, 16.0% to 71.5% and 22.7% to 50.6% respectively, depending on the strength of iterative reconstruction. Potential dose reduction based on clinical image quality varied from 27% to 37.4% depending on the strength of iterative reconstruction. Consequently, the potential dose reduction is strongly dependent on the selected clinical or physical-technical parameter. From the physical-technical image quality parameters, dose reductions based on IQF_{inv} correspond best with dose reductions based on VGAS. The results are summarized in Table 4.

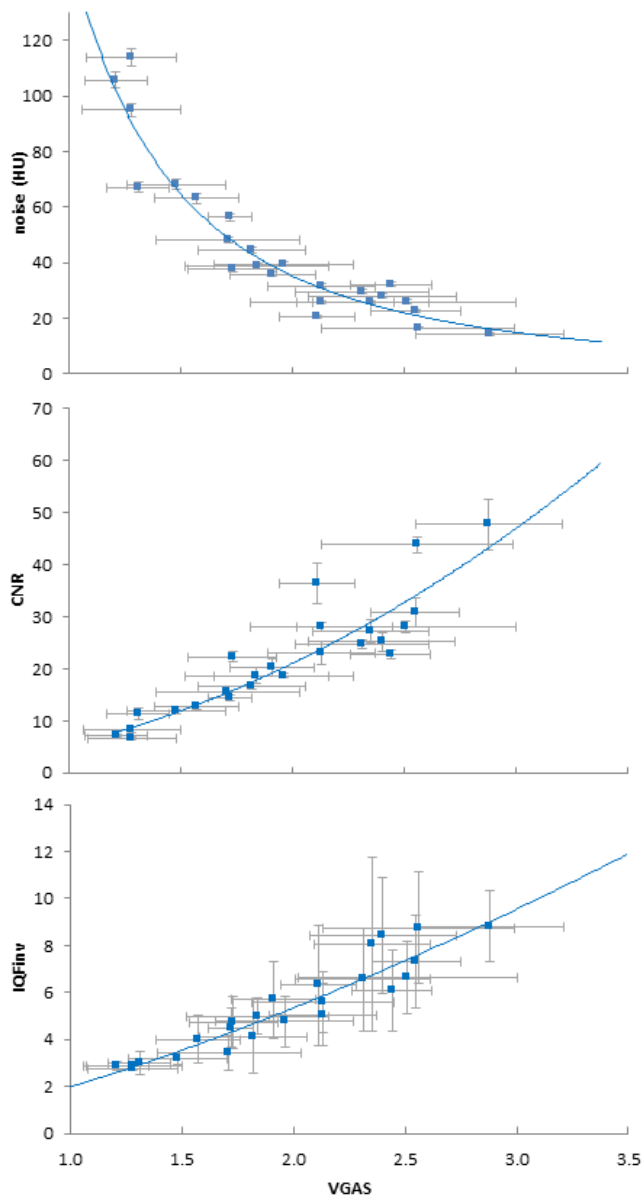


Figure 3: Mean noise, CNR and IQF_{inv} versus mean VGAS. The error bars in the x direction represent the standard deviation between the scores of the different radiologists. For noise, the error bars in the y direction represent the standard deviation between noise measurements in 11 following slices in the Catphan uniformity module. For CNR, the error bars in the y direction represent the standard deviation between CNR measurements in four consecutive slices of the Catphan CT number linearity and CT number accuracy module. For IQF_{inv} the error bars in the y direction represent the standard deviation between the six readers of the contrast-detail module in the Catphan phantom. Regression lines were plotted resulting in an r^2 of 0.90, 0.88 and 0.92, $p < 0.001$, for noise, CNR and IQF_{inv} respectively.

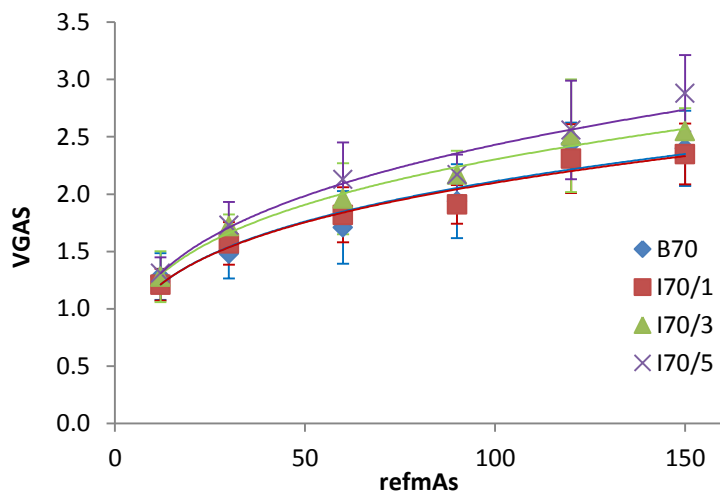


Figure 4: Mean VGAS versus refmAs for B70, I70/1, I70/3 and I70/5. The error bars in the y direction represent the standard deviation between the scores of the different readers. Power functions were plotted and for all kernels an $r^2 > 0.8$ was obtained, $p < 0.05$

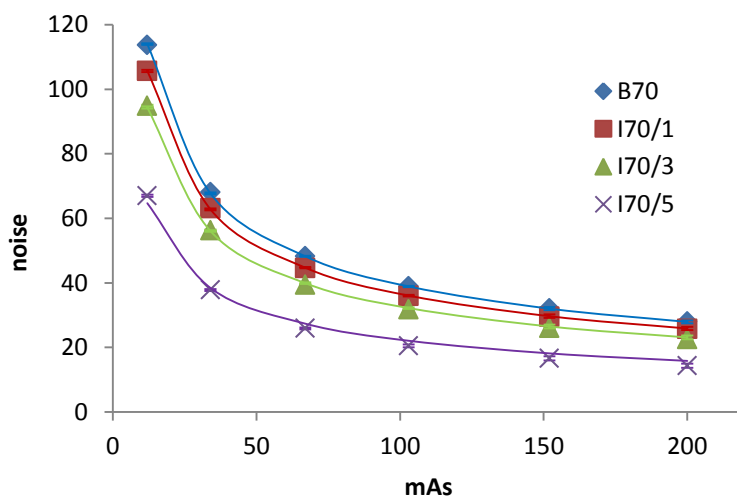


Figure 5: Mean noise versus mAs for B70, I70/1, I70/3 and I70/5. The error bars in the y direction represent the standard deviation between noise measurements in 11 following slices in the Catphan uniformity module. $a/a/mAs$ regression curves were added.

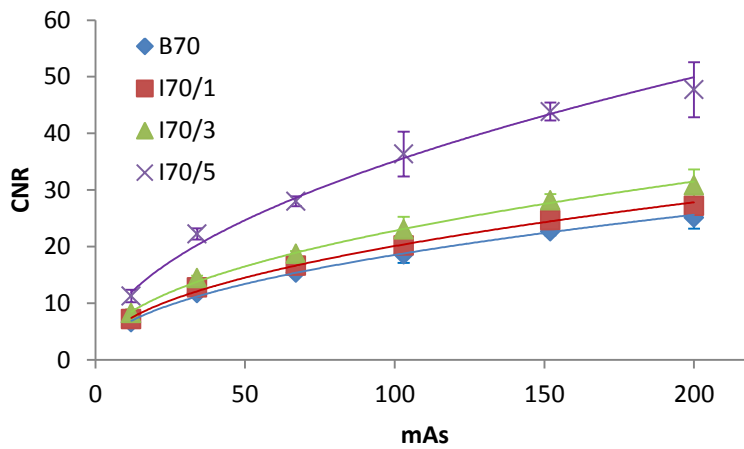


Figure 6: Mean CNR versus mAs for B70, I70/1, I70/3 and I70/5. The error bars in the y direction represent the standard deviation between CNR measurements in four consecutive slices of the Catphan CT number linearity and CT number accuracy module. Power functions were plotted and for all kernels an $r^2 > 0.9$ was obtained, $p < 0.05$.

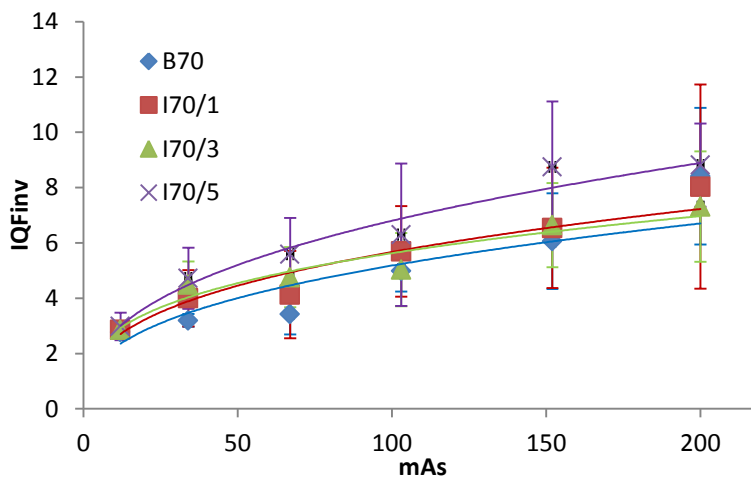


Figure 7: Mean IQF_{inv} versus mAs for B70, I70/1, I70/3 and I70/5. The error bars in the y direction represent the standard deviation between the scores of the different readers. Power functions were plotted and for all kernels an $r^2 > 0.8$ was obtained, $p < 0.05$.

Table 4: Potential dose reduction for different clinical and physical image quality parameters and different iterations steps

Reconstruction kernel	Potential dose reduction (p-value)			
	VGAS	Noise	IQF _{inv}	CNR _{teflon}
l70/1	0% (0.887)	14.0% (< 0.001)	22.7% (0.034)	16.0% (< 0.001)
l70/3	27.0% (0.021)	31.4% (< 0.001)	23.0% (0.031)	35.8% (< 0.001)
l70/5	37.4% (0.001)	67.8% (< 0.001)	50.6% (< 0.001)	71.5% (< 0.001)

DISCUSSION

Methods for patient dose evaluation are easily available but techniques for objective clinical image quality optimization are far more complicated. VGA and ROC studies are commonly used to assess clinical image quality [22]. In VGA studies, a relative or absolute scoring is performed based on the visibility of normal anatomical structures [8]. The task for observers in a ROC study is to detect whether a patient's image contains a pathological structure or not [9]. However, these studies are difficult to implement in routine practice since they imply a significant additional workload for the radiologists and large patient data groups must be available. Therefore, the latter methods are not feasible within a routine quality assurance program.

A more practical approach to assess the image quality is the use of physical-technical phantoms, such as the Catphan phantom, where physical-technical parameters such as noise, CNR, MTF and contrast-detail can be analyzed. Such physical-technical phantoms have been widely used for the objective analysis of the image quality performance of CT systems [23]. Catphan studies are easily implemented in a quality assurance program since no patient data are required and images can be analyzed by the medical physics expert. However, the disadvantage of the Catphan phantom is the uniform background. Actual patient images are clearly not uniform and contain detailed anatomical features and textures. These background anatomical textures can influence image quality, both because the presence of anatomical texture affects observer performance and quantum noise [7].

In present study, a VGA and Catphan study was set up to assess the relationship between physical-technical and clinical image quality in chest CT examinations, using Thiel embalmed cadavers and the Catphan@504 phantom. In contrast to conventional embalming procedures

using formalin for conservation, this new technique results in a very well preservation of the lung structures [11, 12]. To approximate as good as possible the normal patient anatomy, Thiel bodies were ventilated during image acquisition [13]. After assessment of different thoracic regions by experienced radiologists, it was confirmed that Thiel bodies can be applied to assess clinical image quality using VGA and ROC studies.

Recently, there is growing interest in developing and utilizing model observers to accurately predict human observer performance for image system optimization and comparison. A model observer is a mathematical model that can be used to predict human detection performance for some specific imaging tasks [24]. A variety of models, which differ in how much information about signal and noise are used and whether certain properties of the human visual system responses are incorporated, have been proposed and applied to medical image research [24, 25]. However, up till now, phantom images and simulated lesions are used to assess these models. How much real lesions and anatomical backgrounds affect model observer performance remains under investigation [26]. Possibly, the concept of Thiel embalmed cadavers could be used to help validate model observer applications.

Although good correlation was found between physical-technical image quality parameters (noise, CNR and contrast-detail) and clinically observed quality as scored by radiologists (VGAS), the potential dose reduction based on the physical-technical image quality parameters noise and CNR, is much higher compared to the potential dose reduction based on the clinical image quality. This overestimate of the dose reduction can be explained because the uniform phantom does not account for the complex relationship between anatomical variability and image quality. On the contrary, the potential dose reduction based on IQF_{inv} is more consistent with the potential dose reduction based on VGAS. However the measurements are very crude using the Catphan phantom as only three contrast levels are present. Optimization of a contrast-detail phantom for CT is necessary and could give added value to the concept of contrast-detail analysis in CT image quality studies similar to the use of contrast-detail phantoms in mammography and conventional radiology [27, 28].

While this study illustrates that noise measurements in uniform backgrounds are not ideal to assess the effect of iterative reconstruction, a large part of the literature is still based on this technique. Mieville et al. [29] used the Catphan phantom to assess noise, CT number accuracy, noise power spectrum and MTF at varying CTDI values for both FBP images and IR images. Milim Kim et al. [30] used the phantom of the American College of Radiology, a solid water phantom with 5 imbedded test objects to evaluate image noise, SNR and CNR. No

comparisons were made between the possible dose reduction based on the different parameters. Ghetti et al [31, 32] assessed image noise in a uniform water phantom. Since noise reduction in these studies are based on uniform phantoms, it is questionable if these results are applicable in clinical practice. The nonlinear nature of IR methods has also introduced significant challenges to the characterization of spatial resolution performance. In this framework, Li et al introduced a concept of task specific measurements of the spatial resolution by locally measuring the point spread function for a given feature of interest at a given radiation dose level in an anthropomorphic phantom [33].

Other studies performed clinical image quality assessment on patient data, which automatically limits the amount of dose settings that can be used. Exact calculation of the potential dose reduction without loss of image quality is thereby impossible. Prakash et al. [34] scanned 54 patients at a mean effective dose of 12.2 mSv reconstructed with FBP and 98 patients at a mean effective dose of 8.9 mSv reconstructed with an iterative reconstruction technique (30% ASIR, GE) resulting in a mean dose reduction of 27.6%. All chest CT examinations were scored diagnostically acceptable. Pontana et al. [35] scanned 80 patients two times with constant CT parameters except for the refmAs which was decreased by 30%. High dose chest CT images were reconstructed with FBP, low dose chest CT images were reconstructed with an iterative reconstruction technique (IRIS algorithm, Siemens). There was no significant difference in objective noise, CNR, SNR and overall subjective image quality between the two groups. In both studies, physical-technical as well as clinical image quality was assessed. However, no further correlation analysis was made between the physical-technical and clinical image quality. Since only two dose settings were examined, the dose reduction based on clinical and physical-technical image quality was identical. Consequently, the maximum potential dose reduction without loss of image quality, could not be assessed and no conclusions can be made about the discrepancy in potential dose reduction when using physical-technical parameters rather than clinical image quality assessment.

There are several limitations of our study. Firstly, the available Thiel cadavers all had a BMI between 20 and 25. It is possible that the correlation between clinical and physical-technical image quality and the effect of iterative reconstruction can be influenced by patient size. Secondly, clinical image quality was assessed on unenhanced CT images. Possibly the correlation and potential dose reduction can be affected when contrast agents are used. Thirdly, clinical image quality was assessed by a subjective overall quality score and not by means of detection of pathology. Detection of lesions by means of a receiver operating

characteristic (ROC) analysis could give a more precise assessment of image quality for a specific clinical application.

CONCLUSIONS

In summary, our results demonstrate that noise assessments in a uniform phantom overestimate the potential dose reduction for the SAFIRE IR algorithm. Since the IQF_{inv} based dose reduction is quite consistent with the clinical based dose reduction, an optimized contrast-detail phantom could improve the use of contrast-detail analysis for image quality assessment in chest CT imaging. In conclusion, one should be cautious to evaluate the performance of CT equipment taking into account only physical-technical parameters as noise and CNR, as this might give an incomplete representation of the actual clinical image quality performance.

References

1. ICRP, Publication 87: *Managing patient dose in computed tomography*. 2000.
2. Brenner, D.J. and E.J. Hall, *Computed Tomography - An Increasing Source of Radiation Exposure*. The New England Journal of Medicine, 2007. 357: p. 2277-2284.
3. Deak, P.D., et al., *Effects of Adaptive Section Collimation on Patient Radiation Dose in Multisection Spiral CT*. Radiology, 2009. 252(1): p. 140-147.
4. Lee, T.-Y. and R.K. Chhem, *Impact of new technologies on dose reduction in CT*. European Journal of Radiology, 2010. 76: p. 28-35.
5. Kalra, M.K., et al., *Strategies for CT radiation dose optimization*. Radiology, 2004. 230(3): p. 619-28.
6. Kubo, T., et al., *Radiation dose reduction in chest CT: a review*. AJR Am J Roentgenol, 2008. 190(2): p. 335-43.
7. Solomon, J. and E. Samei, *Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE*. Medical Physics, 2014. 41(9): p. 12.
8. Mievile, F.A., et al., *Paediatric cardiac CT examinations: impact of the iterative reconstruction method ASIR on image quality - preliminary findings*. Pediatric Radiology, 2011. 41(9): p. 1154-1164.
9. Martinsen, A.C., et al., *Iterative reconstruction reduces abdominal CT dose*. Eur J Radiol, 2012. 81(7): p. 1483-7.
10. Veldkamp, W., L. Kroft, and J. Geleijns, *Dose and perceived image quality in chest radiography*. European Journal of Radiology, 2009. 72: p. 209-217.
11. Thiel, W., *Die Konservierung ganzer Leichen in natürlichen Farben*. Annals of Anatomy, 1992. 174: p. 185-195.
12. Thiel, W., *Ergänzung für die Konservierung ganze Leichen nach W. Thiel*. Annals of Anatomy, 2002. 184: p. 267-269.

13. De Crop, A., et al., *Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study*. Radiology, 2012. 262(1): p. 298-304.
14. AAPM report no.39, *Specification and acceptance testing of computed tomography scanners*. AAPM, 1993.
15. Samei, E., et al., *Assessment of display performance for medical imaging systems: Executive summary of AAPM TG18 report*. Medical Physics, 2005. 32(4): p. 1205-1225.
16. Commission of the European Communities, *European guidelines on quality criteria for computed tomography* (EUR 16262).
17. Commission of the European Communities, *CEC quality criteria for diagnostic radiographic images and patient exposure trial*. (EUR 12952 EN) 1990.
18. Sund, P., et al., *Comparison of visual grading analysis and determination of detective quantum efficiency for evaluation system performance in digital chest radiography*. European Radiology, 2004. 14: p. 48-58.
19. Borjesson, S., et al., *A software tool for increased efficiency in observer performance studies in radiology*. Radiation Protection Dosimetry, 2005. 114(1-3): p. 45-52.
20. Thijssen, M., K. Bijkerk, and R. van der Burgth, *Manual Contrast-Detail Phantom CDRAD type 2.0. Project Quality Assurance in Radiology*, Department of Radiology, University Hospital Nijmegen, St. Radboud, The Netherlands. 1998.
21. Viner, M., et al., *Liver SULmean at FDG PET/CT: Interreader Agreement and Impact of Placement of Volume of Interest*. Radiology, 2013. 267(2): p. 596-601.
22. Bath, M., *Evaluating imaging systems: practical applications*. Radiat Prot Dosimetry, 2010. 139(1-3): p. 26-36.
23. Baker, M.E., et al., *Contrast-to-Noise Ratio and Low-Contrast Object Resolution on Full- and Low-Dose MDCT: SAFIRE Versus Filtered Back Projection in a Low-Contrast Object Phantom and in the Liver*. American Journal of Roentgenology, 2012. 199(1): p. 8-18.
24. Barrett, H.H., et al., *Model Observers for Assessment of Image Quality*. Proceedings of the National Academy of Sciences of the United States of America, 1993. 90(21): p. 9758-9765.
25. Richard, S. and J.H. Siewerdsen, *Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images*. Medical Physics, 2008. 35(11): p. 5043-5053.
26. Zhang, Y., et al., *Correlation between human and model observer performance for discrimination task in CT*. Physics in Medicine and Biology, 2014. 59(13): p. 3389-3404.
27. Yip, M., et al., *Validation of a Simulated Dose Reduction Methodology Using Digital Mammography CDMAM Images and Mastectomy Images*. Digital Mammography, 2010. 6136: p. 78-85.
28. Veldkamp, W., et al., *Contrast-detail evaluation and dose assessment of eight digital chest radiography systems in clinical practice*. European Radiology, 2006. 16: p. 333-341.
29. Mievile, F.A., et al., *Iterative reconstruction methods in two different MDCT scanners: Physical metrics and 4-alternative forced-choice detectability experiments - A phantom approach*. Physica Medica-European Journal of Medical Physics, 2013. 29(1): p. 99-110.
30. Kim, M., et al., *Adaptive Iterative Dose Reduction Algorithm in CT: Effect on Image Quality Compared with Filtered Back Projection in Body Phantoms of Different Sizes*. Korean Journal of Radiology, 2014. 15(2): p. 195-204.
31. Ghetti, C., O. Ortenzia, and G. Serreli, *CT iterative reconstruction in image space: A phantom study*. Physica Medica-European Journal of Medical Physics, 2012. 28(2): p. 161-165.

32. Ghetti, C., et al., *Physical characterization of a new CT iterative reconstruction method operating in sinogram space*. Journal of Applied Clinical Medical Physics, 2013. 14(4): p. 263-271.
33. Li, K., et al., *Statistical model based iterative reconstruction (MBIR) in clinical CT systems. Part II. Experimental assessment of spatial resolution performance*. Medical Physics, 2014. 41(7): p. 12.
34. Prakash, P., et al., *Radiation Dose Reduction With Chest Computed Tomography Using Adaptive Statistical Iterative Reconstruction Technique: Initial Experience*. Journal of Computer Assisted Tomography, 2010. 34(1): p. 40-45.
35. Pontana, F., et al., *Chest computed tomography using iterative reconstruction vs filtered back projection (Part 2): image quality of low-dose CT examinations in 80 patients*. European Radiology, 2011. 21(3): p. 636-643.

Analysis of metal artifact reduction tools for dental hardware in CT scans of the oral cavity: kVp, Iterative Reconstruction, Dual Energy CT, Metal Artifact Reduction Software: Does it make a difference?

7.1 Background

7.1.1 Dual energy CT

The term dual-energy CT or DECT refers to CT that uses two photon spectra; therefore, DECT is sometimes also referred to as spectral CT. In clinical practice today, two different spectra are generated either by switching the voltage of one X-ray tube or by running two tubes at different voltages. The dual source system consists of 2 X-ray tubes that acquire the data with different energies in a DE acquisition. One of the advantages of this technique is the availability of tube current modulations for optimizing radiation dose. However, one of the limitations of this system is the slight difference in the acquisition time of the 2 data sets, resulting in a limited temporal registration [21]. The single source dual energy system uses a single X-ray tube, which can alternate tube energy in less than 0.5 ms, in the same gantry rotation. It also relies on a detector with much faster response that is able to register information from both acquisitions. It has a better temporal resolutions as both data sets are

acquired near simultaneously. Nevertheless, the use of tube current modulations is not supported in this system, which can result in some increased radiation dose [22].

Although the term dual-energy CT suggests that two distinct photon energies are used, it rather are two X-ray spectra. The settings of 80 and 140kVp are commonly used because they provide the maximum difference and least overlap between the spectra with standard tubes [22].

The 2 different energy settings allow for the differentiation of materials based on their energy related attenuation characteristics. In general, the attenuation of material represented by its CT number is caused by a combination of photo-electric and Compton effects. These two main mechanisms contributing to CT attenuation are both energy- and material dependent. The probability of an X-ray undergoing photoelectric effect increases in substances with higher atomic number and is heavily energy dependent with an increase at lower energy levels. Compton scattering occurs almost independently of the photon energy at energies exceeding 30keV and is predominantly related to the density of the material. Because the photo-electric effect is energy dependent and linked to the atomic number, it is possible to derive information about a given element from the attenuation at different energy levels. If there is a difference in the photo-electric behavior of two elements, their attenuation at two distinct energies is different, thus allowing for separation of the elements from each other. This concept is used in dual energy CT [21, 22].

Various clinically relevant applications have been established for dual energy CT based on either differentiation or quantification of materials. By differentiating calcium and iodine, bone removal can be performed automatically which can be an interesting solution in complex anatomic areas with a close proximity of vascular and bony structures such as the skull [23, 24]. The detection of iodine in CT datasets allows for assessment of tissue enhancement which gives information about lesion vascularity [25]. Dual energy derived iodine distribution maps also allows accurate assessment of cardiac perfusion defects [26].

The detection and characterization of iodine in CT datasets not only allows for its quantification but also for its subtraction. By subtracting iodinated contrast material from contrast enhanced images, a virtual non-enhanced image is generated. This application lowers the radiation dose since non-enhanced acquisitions can be omitted. Virtual non-enhanced images were shown to be an acceptable substitute for standard non-enhanced acquisition for different clinical indications such as the detection of urinary stone disease [27], liver imaging [28] and intracranial hemorrhage [29].

Another application of dual-energy CT is to generate virtual monochromatic images from the dual energy CT scans. In clinical practice, all X-ray beams generated by single energy CT are polychromatic. Virtual monochromatic images depict objects as if they were theoretically images with a monochromatic beam. The X-ray energy is then reported as keV instead of kilopeak voltage (kVp). These monochromatic images have the potential to reduce beam-hardening artifacts [30-32] and improve image quality in a different range of energies (from 40-140keV) that can be generated. The selection of the virtual monochromatic energy should be tuned to a diagnostic task.

The radiation exposure required for DECT depends on the technology used. Generally, the aim is to use the same dose as would be used for a single energy examination. Only then, it is easily possible to replace standard protocols with dual energy examinations because an additional diagnostic value is offered without additional dose. The current second generation dual source DECT system is almost dose neutral to single energy scanning [33]. This technological improvement has been attributed to the use of a tin filter on the 140 kV source, which also enables better material differentiation and tissue contrast [21]. Moreover, the ability to apply iterative reconstruction to dual-energy raw data enables further image quality refinement [34, 35]. The single source DECT system with rapid kVp switching has also been recently upgraded to a second generation scanner with a concurrent iterative reconstruction capability and has the potential to achieve dose neutral scans [21].

7.1.2 Metal artifacts

In CT, the term artifact is applied to any systematic discrepancy between the CT numbers in the reconstructed image and the true attenuation coefficients of the object. The presence of metal in an object can be the cause of such artifacts, in this case called: metal artifacts. These artifacts can be seen as light or dark bands or streaks in the reconstructed image.

Metal artifacts originate from two main processes: beam hardening and photon starvation. Beam hardening artifacts originate from the polychromaticity of an X-ray beam. When an X-ray beam passes through an object, the lower energy photons are absorbed more rapidly than the higher energy photons resulting in a mean energy increase [36]. Photon starvation occurs when an X-ray beam is completely absorbed by an object and the number of photons reaching the X-ray detector is insufficient for correct image reconstruction [30, 36]. These processes can

seriously deteriorate the image quality of a CT scan, thereby interfering with the diagnostic interpretability.

Patients are normally asked to take off removable metal objects such as jewelry before scanning commences. However, when non removable items such as prosthetic devices, surgical clips and dental fillings are present, corrective actions might be needed to compensate for the induced metal artifacts.

7.1.3 Head and neck CT

In 2011, the frequency of head and neck CT examinations in Belgium was 50 and 1 per 1000 of population respectively [4]. In the same period, the average amount of head and neck CT examinations over 36 European countries was 27 and 4 per 1000 of population respectively [4]. In 2013, about 68 000 neck CT examinations, i.e. 3.3% of all CT examinations, were performed in Belgium [5].

The mean effective dose per patient estimated in European countries for head and neck CT examinations is 1.9 and 2.5 mSv respectively, in Belgium this is 1.3 and 2.9 mSv respectively [4] (Figure 7-1, Figure 7-2).

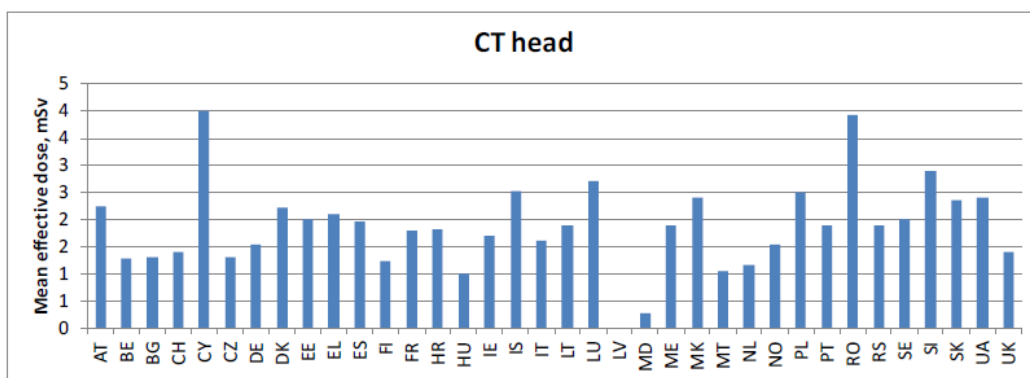


Figure 7-1: Typical effective doses (mSv) for head CT in different European countries [4]

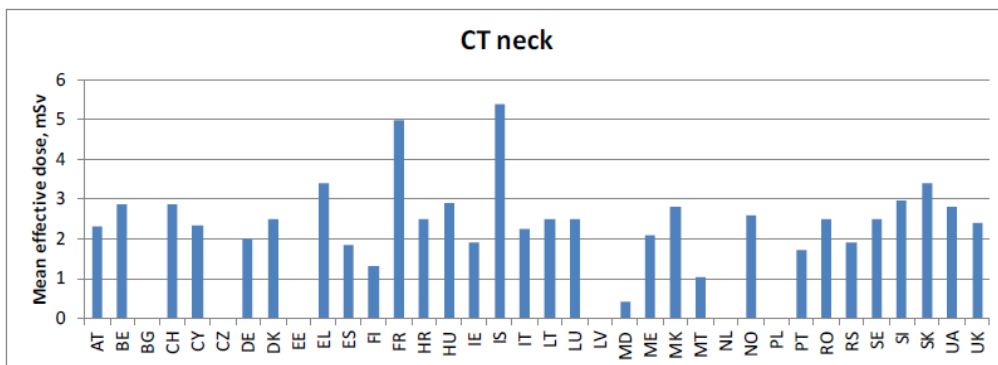


Figure 7-2: Typical effective doses (mSv) for neck CT in different European countries [4]

7.2 Paper III

Analysis of metal artifact reduction tools for dental hardware in CT scans of the oral cavity: kVp, Iterative Reconstruction, Dual Energy CT, Metal Artifact Reduction Software: Does it make a difference?

An De Crop¹, Jan Casselman², Tom Van Hoof¹, Melissa Dierens³, Elke Vereecke², Nicolas Bossu², Jaime Pamplona⁴, Katharina D'Herde¹, Hubert Thierens¹, Klaus Bacher¹

¹Ghent University, Department of Basic Medical Sciences, Proeftuinstraat 86, B-9000 Gent

²AZ Sint Jan Bruges Ostend AV, Department of Radiology, Ruddershove 10, B-8000 Bruges

³Ghent University, Dental School, Unit for Oral and Maxillofacial Imaging, De Pintelaan 185 P8, B-9000 Ghent

⁴Hospital Lisboa Central, Department of Neuroradiology, 1169 Lisbon, Portugal

Reprint from Neuroradiology: 2015; epub ahead of print

ABSTRACT

Introduction: Metal artifacts may negatively affect radiologic assessment in the oral cavity. The aim of this study was to evaluate different metal artifact reduction techniques for metal artifacts induced by dental hardware in CT scans of the oral cavity.

Methods: Clinical image quality was assessed using a Thiel embalmed cadaver. A Catphan phantom and a PMMA phantom were used to evaluate physical-technical image quality

parameters such as artifact area, artifact index (AI) and contrast detail (IQF_{inv}). Metal cylinders were inserted in each phantom to create metal artifacts. CT images of both phantoms and the Thiel embalmed cadaver were acquired on a multislice CT scanner using 80, 100, 120 and 140 kVp; model-based iterative reconstruction (Veo) and synthesized monochromatic keV images with and without metal artifact reduction software (MARs). Four radiologists assessed the clinical image quality, using an image criteria score (ICS).

Results: Significant influence of increasing kVp and the use of Veo was found on clinical image quality ($p = 0.007$ and $p = 0.014$, respectively). Application of MARs resulted in a smaller artifact area ($p < 0.05$). However, MARs reconstructed images resulted in lower ICS.

Conclusion: Of all investigated techniques, Veo shows to be most promising, with a significant improvement of both the clinical and physical-technical image quality without adversely affecting contrast detail. MARs reconstruction in CT images of the oral cavity to reduce dental hardware metallic artifacts is not sufficient, and may even adversely influence the image quality.

INTRODUCTION

CT plays a very important role in both diagnosis and treatment of head and neck cancers. It is a standard method for identifying cancerous masses, lymph node metastases and inflammatory processes in the head and neck region [1]. For both the radiologist and the surgeon, the information provided by the images enables accurate tumor staging and treatment planning prior to surgery. The image quality and therefore suitability for accurate diagnoses can be significantly reduced by dental artifacts. CT also plays an important role in radiotherapy planning. Radiotherapy critically relies on the ability to precisely delineate target volumes. As soft tissue contrast is limited, the image quality of the planning CT is often a limiting factor in the exact determination of boundaries. The visual reduction of metal artifacts therefore becomes very important.

The main fundamental causes of metal artifacts are beam hardening and photon starvation [2]. Beam hardening is the absorption of low-energy photons in a polychromatic X-ray beam when passing through metal, which leads to an increase in the average energy of the beam, resulting in dark streaks on the image. Photon starvation occurs as the metal absorbs a large proportion of photons which also creates streaks in the image.

The factors that may contribute to or diminish the artifacts include metallic hardware composition and orientation [3-5], peak voltage[6-8], slice thickness [6], reconstruction algorithm [3, 9] and extended CT scale [10]. Over the years, many metal artifact reduction algorithms have been proposed for the suppression of streak artifacts [8, 11]. Another possibility is the use of iterative reconstruction algorithms [8]. Essentially, iterative reconstruction introduces a correction loop in the image generation process that cleans up artifacts and noise. GE (GE Medical Systems) introduced a full iterative reconstruction or model-based iterative reconstruction (MBIR- under the commercial name of VEO). VEO is a fully iterative method working in the raw data domain, which takes not only the data statistics into account but also the geometry of the CT equipment itself. Since beam hardening artifacts originate from the polychromaticity of the X-ray beam, monochromatic images could reduce metal artifacts. Recent development in dual-energy CT (DECT) provided the ability to generate monochromatic images [12]. Consequently, DECT combined with monochromatic image reconstruction has been investigated to reduce metal artifacts [13-15].

The aim of this study was to evaluate different metal artifact reduction techniques for metal artifacts induced by dental hardware in CT scans of the oral cavity. Moreover, the best of each technique were compared in relations to standard 120 kVp filtered back projection (FBP) imaging to define the optimal metal artifact reduction technique for this particular CT examination.

MATERIALS AND METHODS

Phantoms and inserted metals

Three different materials were used to examine the amount of induced metal artifacts: Cobalt-Chromium (CoCr), Titanium (Ti) and Zirconium (Zr). Ti is commonly used for dental implants because of its excellent biocompatibility with bone tissue. The implant is usually cylindrical with a diameter varying from 3 to 8 mm and a length of 6 to 18 mm. Ti, CoCr and Zr are used as the core of prosthetic suprastructures, the geometry of these structures is adapted to the individual needs of the patient. The metals were fabricated in 4mm cylinders with a length of 10mm (Proscan, Zonhoven, Belgium). For the technical evaluation of the image quality, these cylinders were inserted in a 16 cm homogenous cylindrical polymethylmethacrylate (PMMA) phantom. Two 4mm holes (one on the left and one on the right) were drilled at the periphery of the phantom for the insertion of the metal cylinders. To avoid the influence of air around

the metal cylinders, PMMA was placed at the front and the back of the metal cylinder, allowing the insertion to be completely filled with metal and PMMA.

To evaluate contrast detail, the Catphan@504 phantom (The Phantom laboratory, Salem, New York, USA) was used. The phantom consists of several modules to evaluate high and low contrast resolution, CNR and noise. In the low contrast module there are three areas with different contrast levels: 1%, 0.5% and 0.3%. Each contrast level contains targets with decreasing diameters (15, 9, 8, 7, 6, 5, 4, 3 and 2 mm). A CoCr cylinder was placed at the level of the 0.3% contrast level on the outside of the phantom to create metal artifacts (Fig 1).

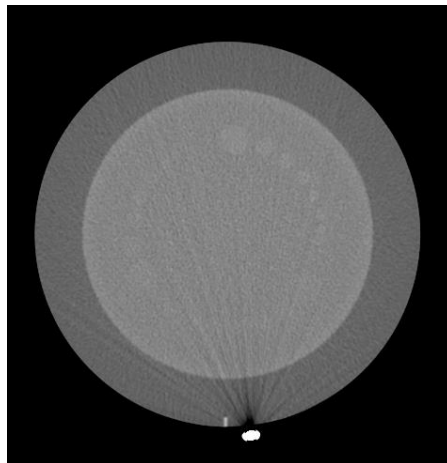


Fig 1: Low contrast module of the Catphan@504 phantom. Three areas with different contrast levels (1%, 0.5% and 0.3%) are present. Each contrast level contains targets with decreasing diameters (15, 9, 8, 7, 6, 5, 4, 3 and 2 mm). A CoCr cylinder was placed at the level of the 0.3% contrast level on the outside of the phantom to create metal artifacts

A Thiel embalmed cadaver was used for the clinical evaluation of the image quality. The use of a human cadaver fulfilled the requirements of the ethical committee of our institution. One human cadaver was embalmed using the methodology of Prof. Em. Walther Thiel, Anatomisches Institut Karl-Franzens-Universität, Graz, Austria [16].

Hereby, 4-chloro-3-methylenphenol as well as various salts are used for fixation and boric acid is added for disinfection. Furthermore, ethylene glycol is used for preservation of tissue plasticity, while the concentration of formalin is kept to the strict minimum (0.8%) [17]. In contrast to standard formalin-embalmed human cadavers, this technique results in well preserved organs and tissues concerning color, consistency, natural flexibility and natural plasticity. Two 4 mm holes were drilled in the maxilla to obtain the same configuration as in

the physical-technical measurements in the PMMA phantom, i.e. 2 inserts placed towards the molars on the left and right side of the maxilla.

Image acquisition

All examinations were performed using a clinical 64 slice CT system (GE Discovery CT 750 HD, GE Medical Systems) with fast tube voltage switching between two energies (80 kVp and 140 kVp) within 0.5 ms. Two different scan protocols were used.

First, all phantoms were scanned in helical acquisition mode, beam collimation 20 mm, scanning field of view 32 cm, reconstruction field of view 25 cm, pitch 0.963:1, slice thickness 2.5 mm, slice increment 2 mm, rotation speed 1 s. To study the effect of kVp on the amount of metal artifacts, scans were acquired using 80, 100, 120 and 140 kVp with 580, 320, 210 and 150 mAs, respectively, keeping the CT dose index (CTDI) constant at 40 mGy. To investigate the influence of iterative reconstruction (IR), the 120 kVp acquisitions were reconstructed with FBP and with model based iterative reconstruction (Veo). To examine the effect of kVp and IR, all scans were performed using CoCr insertions. However, to examine the amount of artifacts caused by different materials, two 120 kVp, FBP scan using Zr and Ti insertions were acquired.

Second, the GSI-MARs protocol 20 was selected, beam collimation 20 mm, scanning field of view 32 cm, reconstruction field of view 25 cm, pitch 0.936:1, slices thickness 2.5 mm, slices increment 2 mm, rotation speed 0.5 s, current less than 630 mA with a CTDI of 44.61 mGy. All images were reconstructed both with and without metal artifact reduction software (MARs). Post-processing was applied to generate synthesized monochromatic 40-140 keV images with 10 keV interval of each image series by using dedicated GSI viewer software (GSI viewer 2.00 and GE VolumeShare4 AW 4.4, GE Healthcare). CoCr insertions were used to examine the effect of keV and MARs.

For both scan protocols, reference scans were made of the PMMA phantom in the absence of any metal cylinder using identical reconstruction settings as in the scans with metal insertions.

Image quality analysis

- PMMA phantom

A doughnut shaped region of interest (ROI) with a band size of 5 mm was defined around the left metal insertion. According to Lin et al [18] and Wang et al [14], the artifact index (AI) to quantify the severity of metal artifacts is defined as $\sqrt{SD_o - SD_b}$, where SD_o is the noise of the doughnut shaped ROI in the investigated series and SD_b the noise of the same ROI in the corresponding reference scan.

Based on van der Schaaf et al [19], artifact areas are quantitatively analyzed by establishing threshold attenuation values for black and white artifacts. To examine the influence of the selected kVp, threshold values are calculated by taking the attenuation value in the corresponding reference scan and adding or subtracting three times the standard deviation of the reference scan with the lowest theoretical exposure parameters (80 kVp) as this dataset was judged to have the maximum amount of noise. To analyze the influence of the selected keV and degree of iterative reconstruction, threshold values are calculated similar as before, using the standard deviation of the 40 keV reference scan and the 120 kVp, FBP reference scan respectively. The artifact areas in the ROI covering the entire phantom were quantified in millimeters squared. The total area was calculated as the sum of the black and white areas. Two areas, representing the area of the metal insertion itself, were subtracted in each image.

- *Catphan phantom*

Four medical physicists identified the minimally visible target diameter at three different contrast levels. The inverse image quality figure (IQF_{inv}) was introduced for quantitative comparison of the contrast-detail images [20]. The inverse image quality figure is defined as

$$IQF_{inv} = \frac{100}{\sum_{i=1}^n C_i D_{i,th}}$$

where $D_{i,th}$ denotes the threshold diameter for contrast i , and C_i denotes the contrast value. The higher the IQF_{inv} , the better the low-contrast visibility. The IQF_{inv} was calculated for all analyzed images and averaged over the four readers. Moreover, for all images, the threshold diameter for the 0.3% contrast level was analyzed.

- *Thiel embalmed cadaver*

After acquisition, all data were displayed on a 20-inch, 2-megapixel gray-scale monitor (Barco MFGD 2320, Kortrijk, Belgium). The monitor was calibrated to comply with the DICOM Part 3.14 Greyscale Standard Display Function, using calibration software provided by the manufacturer (MediCal Pro, BARCO, Kortrijk, Belgium) [21]. Maximum luminance of all monitors was adjusted to 320cd/m² and ambient lighting levels were below 50 lux as recommended by AAPM TG 18 [22].

Four experienced radiologists (JC: 31 years of experience; EV and NB: 2 years of experience and JP: 6 years of experience) assessed the head CT scans and scored the image quality using the criteria listed in Table 1. Each stack was compared with the reference stack, the 120 kVp

FBP scan, as this is the clinically used setting. Each structure was rated on a scale from -3 to +3 according to Table 2. An image criteria score (ICS) for each reader was calculated as:

$$ICS = \sum_{c=1}^C S_c$$

where S_c is the rating for a particular criteria and C is the number of criteria.

Table 1: Image quality criteria for CT scan of the oral cavity

Criterion No.	Description:
1	Left masseter
2	Right masseter
3	Platysma
4	Cervical spinal cord
5	Tail of the parotid gland
6	Attachment of the medial pterygoid muscle
7	Attachment of the buccinator muscle on the alveolar ridge
8	Tongue base

All series were evaluated by the radiologists using Viewdex, a Java-based DICOM-compatible software tool for presentation and evaluation of images without influencing the image quality. The readers were allowed to adjust the image brightness and contrast and to magnify the images to full resolution. The image stacks were presented in random order. Before starting the study, a training session was organized to familiarize the readers with the scoring methodology.

Table 2: Rating used to evaluate the clinical images

Rating	The structure in the image is:
-3	Far inferior
-2	Noticeable worse
-1	Little worse
0	Equal
+1	Little superior
+2	Noticeable superior
+3	Far superior

Statistical analysis

Inter-observer agreement for ICS, IQF_{inv} and low contrast diameter values was determined by calculating the intraclass correlation coefficient (ICC). An intraclass correlation coefficient greater than 0.9 was considered to suggest an excellent inter-observer agreement [23, 24]. To determine the influence of different exposure and reconstruction settings, data were analyzed using the Kruskal Wallis test, a signed rank, non-parametric test used when comparing more than two independent samples. When only two independent samples needed to be compared, the Mann Whitney U test, a signed rank, non-parametric test, was used.

A 95% confidence interval was used for all statistical measures. All calculations were performed in the SPSS software tool (IBM SPSS statistics 22, IBM corp., NY, USA).

RESULTS

Excellent inter-observer agreement among the participating radiologists and among medical physicists was found by means of an intraclass correlation coefficient of 0.957 ($p < 0.001$), 0.932 ($p < 0.001$) and 0.925 ($p < 0.001$) for VGAS, IQF_{inv} and threshold low contrast diameter respectively.

The influence of the inserted material was investigated for 120 kVp FBP scans by analyzing ICS, artifact area and AI. For all parameters, a significant influence of the inserted material could be found ($p < 0.05$) (Table 3).

Table 3: Clinical and physical-technical image quality for different metal insertions

Material	ICS (\pm SD)	Artifact area (\pm SD)	AI (\pm SD)
CoCr	0.0 \pm 0.0	2413.6 \pm 531.8	5.7 \pm 1.4
Ti	13.7 \pm 3.3	606.7 \pm 168.4	2.3 \pm 0.2
Zr	-5.2 \pm 4.3	3716.5 \pm 416.3	7.0 \pm 0.8

Note: SD = standard deviation; ICS = image criteria score; AI = artifact index; CoCr = Cobalt-Chromium; Ti = Titanium ; Zr = Zirconium.

An increase in kVp, while keeping the CTDI_{vol} constant, resulted in an improvement of the clinical image quality ($p = 0.007$) (Fig 2). Analysis of the physical-technical image quality parameters resulted in a significant improvement for artifact area and AI ($p < 0.05$). Though a significant increase of threshold low contrast diameter could be seen when increasing the kVp ($p < 0.05$). No significant influence could be found for IQF_{inv} ($p = 0.154$).

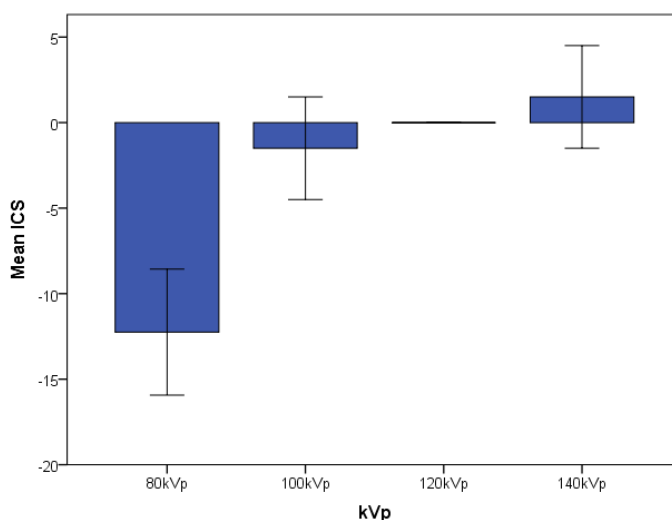


Fig 2: ICS as a function of kVp. The error bars represent the standard deviation between the scores of the different radiologists. A significant influence of the kVp was found ($p = 0.007$). ICS = image criteria score

Application of Veo resulted in a significant improvement for ICS (Table 4, Figure 3), artifact area, IQF_{inv} and threshold low contrast diameter ($p < 0.05$). A clear decrease of the AI could be observed when using Veo (Table 4), though this was not significant ($p = 0.248$).



Fig 3: Thiel head CT scan reconstructed with Veo. Veo decreases the induced metal artifacts resulting in an increased ICS. ICS = image criteria score

¹ Comparison between a 120kVp FBP reconstruction and a Veo reconstruction is displayed in the addendum on page 107

Table 4: Clinical and physical-technical image quality for the comparison of different metal artifact reduction tools with the clinically applied 120kVp FBP protocol

	120kVp FBP	140kVp FBP	120kVp Veo	140keV without MARs
ICS (\pm SD)	0.0 \pm 0.0	1.5 \pm 3.0 (p = 0.316)	5.2 \pm 2.7 (p = 0.014)	5.5 \pm 3.8 (p = 0.046)
Artifact area (\pm SD)	2413.6 \pm 531.8	4504.3 \pm 636.7 (p = 0.043)	1880.9 \pm 162.8 (p = 0.021)	731.6 \pm 101.4 (p = 0.021)
AI (\pm SD)	5.7 \pm 1.4	6.9 \pm 1.1 (p = 0.248)	4.7 \pm 0.4 (p = 0.248)	3.2 \pm 0.5 (p = 0.043)
IQF _{inv} (\pm SD)	0.12 \pm 0.01	0.12 \pm 0.02 (p = 0.877)	0.17 \pm 0.03 (p = 0.058)	0.06 \pm 0.01 (p = 0.019)
Threshold low contrast diameter (\pm SD)	9.5 \pm 3.7	7.2 \pm 1.2 (p = 0.234)	5.7 \pm 0.5 (p = 0.017)	16 \pm 5.2 (p = 0.056)

Note: SD = standard deviation; ICS = image criteria score; AI = artifact index; IQF_{inv} = inverse image quality figure; FBP = filtered back projection; MARs = metal artifact reduction software.

A significant influence of keV was seen for the clinical image quality when images were reconstructed without MARs ($p < 0.001$). For MARs reconstructed images, a significant influence of keV could not be confirmed ($p = 0.099$), however, an increasing trend of ICS could be seen when increasing the keV (Fig 4). For all physical-technical image quality parameters, a significant influence of keV could be confirmed for both MARs and standard reconstructed images ($p < 0.05$), except for the AI in MARs reconstructed images ($p = 0.058$).

The use of MARs did not give the expected results, as overall, MARs reconstructed images resulted in a lower ICS than images reconstructed without MARs (Fig 4). This negative effect is most pronounced for higher keV values, though no significant difference was found for 90 and 110 keV ($p > 0.05$). For lower keV values (40-60 keV), MARs seems not to have any significant effect ($p > 0.05$). These clinical results are partly supported by the results of the contrast detail analysis. For low keV values, MARs reconstructed images result in a higher IQF_{inv} value or a lower threshold low contrast diameter, while for higher keV values, MARs reconstructed images result in lower IQF_{inv} values or a higher threshold low contrast diameter compared to images reconstructed without MARs (Fig 5 and 6).

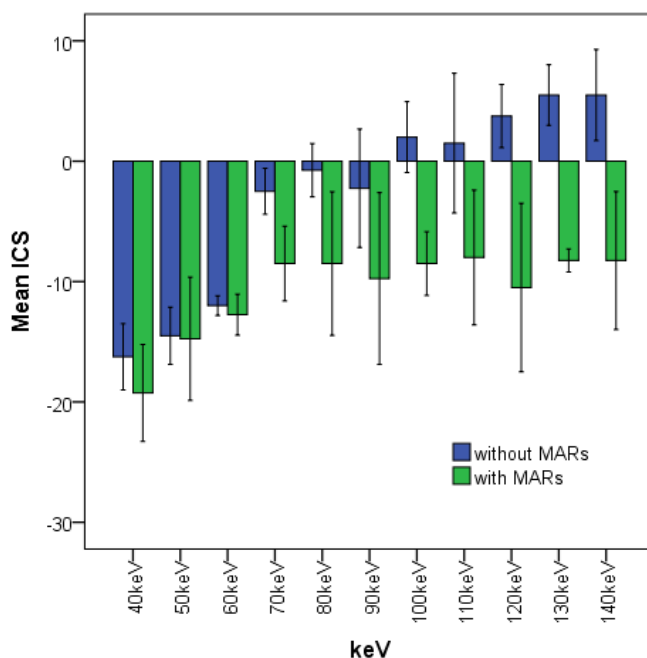


Fig 4: ICS as a function of keV for both MARs and standard synthesized monochromatic images. The error bars represent the standard deviation among the scores of the different radiologists. Significant influence of keV for standard monochromatic images was found ($p < 0.001$). An increasing trend of ICS in MARs reconstructed images could be seen when increasing the keV, though, no significance could be demonstrated ($p = 0.099$). Overall, MARs reconstructed images resulted in a lower ICS than images reconstructed without MARs, though this negative effect was more pronounced for higher keV values. ICS = image criteria score; MARs = metal artifact reduction software

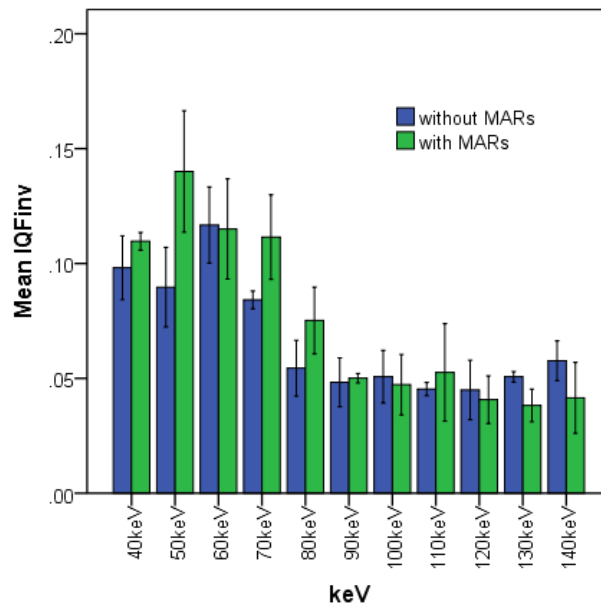


Fig 5: IQF_{inv} as a function of keV for both MARs and standard synthesized monochromatic images. The error bars represent the standard deviation among the scores of the different readers. Significant influence of keV on IQF_{inv} for images reconstructed with and without MARs was found ($p < 0.001$). For low keV values, MARs reconstructed images resulted in a higher IQF_{inv} value, while for higher keV values, MARs reconstructed images resulted in lower IQF_{inv} values, however no significance could be demonstrated ($p > 0.05$). IQF_{inv} = inverse image quality figure; MARs = metal artifact reduction software

However, large variations of IQF_{inv} and threshold low contrast diameter were observed, so overall, no significant influence of MARs could be demonstrated. For all keV values, the noise based physical-technical image quality parameters (artifact area (Fig 7) and AI (Fig 8)) did show a significant improvement when applying MARs ($p = 0.021$).

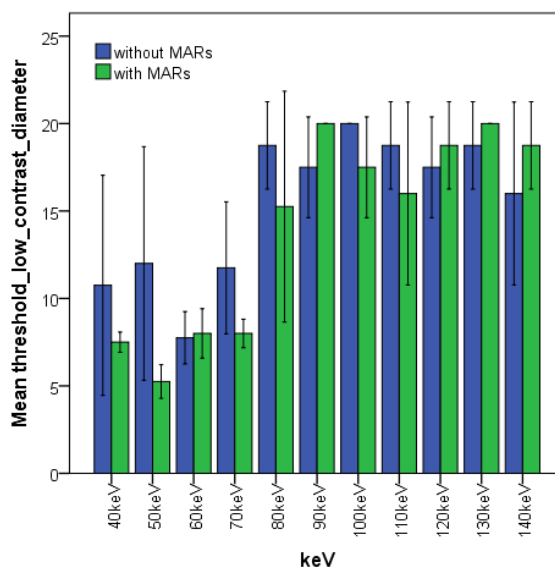


Fig 6: Threshold low contrast diameter as a function of keV for both MARs and standard synthesized monochromatic images. The error bars represent the standard deviation among the scores of the different readers. Significant influence of keV on the threshold low contrast diameter for images reconstructed with and without MARs was found ($p < 0.001$ and $p = 0.012$ respectively). For low keV values, MARs reconstructed images resulted in a lower threshold diameter, while for the highest keV values, MARs reconstructed images resulted in higher low contrast diameters. However no significance could be demonstrated ($p > 0.05$). MARs = metal artifact reduction software

To compare the effect of the different analyzed metal artifact reduction techniques, the series with the optimal clinical and physical-technical image quality within each investigated technique were selected. Subsequently, comparison was made between these selected series and the clinically applied 120 kVp FBP protocol to select the most appropriate scanning/reconstruction technique to reduce metal artifacts from dental hardware in CT images of the oral cavity (Table 4).

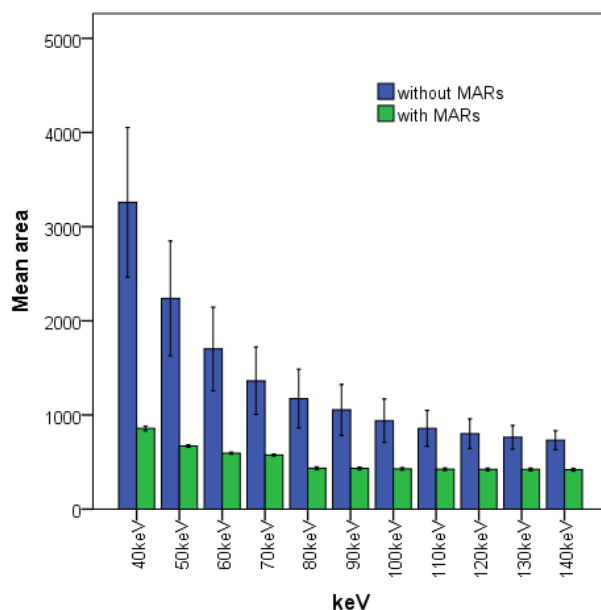


Fig 7: Artifact area as a function of keV for both MARs and standard synthesized monochromatic images. The error bars represent the standard deviation among area measurements in four consecutive slices in the PMMA phantom. A clear decrease of the area is observed when increasing the keV for both images reconstructed with and without MARs ($p < 0.001$). For all keV values, the area showed a significant improvement when applying MARs ($p = 0.021$). MARs = metal artifact reduction software

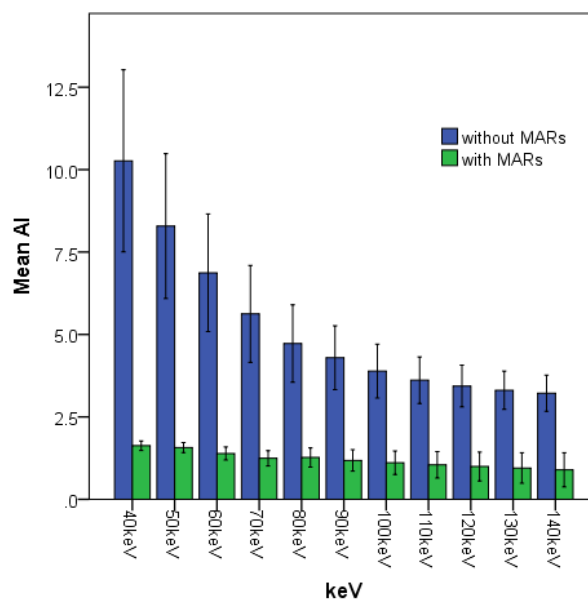


Fig 8: AI as a function of keV for both MARs and standard synthesized monochromatic images. The error bars represent the standard deviation among AI measurements in four consecutive slices in the PMMA phantom. A clear decrease of the AI is observed when increasing the keV for images reconstructed without MARs ($p < 0.001$). For MARs reconstructed images, no significant influence of the keV was found ($p = 0.058$). For all keV values, the area showed a significant improvement when applying MARs ($p = 0.021$). AI = artifact index; MARs = metal artifact reduction software; PMMA = polymethylmethacrylate

DISCUSSION

In the diagnosis of oral cancer, multi-detector CT is a first-line diagnostic device because of its broad availability, the ability to perform whole-body tumor staging, and an overall good sensitivity and specificity for the detection of oral cancer [1]. Consequently, excellent image quality is essential. However, due to the complex anatomy in the oral cavity, the presence of tissues and materials with a wide range of CT numbers and the metal artifacts caused by dental hardware, imaging of the oral cavity is challenging [25]. Metal artifacts induced by high attenuation objects consist of hypodense or hyperdense streaks or radially emerging dark and bright bands on CT images [2]. Such streak artifacts are a common finding on standard oral cavity CT images and may negatively affect the radiologic assessment.

The ultimate solution is to avoid the metal artifacts by using low-level attenuating materials. Many studies have shown that titanium prostheses cause fewer artifacts on CT than CoCr [4, 5, 26]. In this study, analysis of the clinical and physical-technical image quality confirmed that Ti causes the least amount of metal artifacts or the highest ICS. CoCr clearly results in more

metal artifacts and a lower ICS although it is the recently more frequently applied material Zr that causes the most artifacts or the lowest ICS (Table 3).

This study demonstrated that for the same dose level, higher tube voltage was associated with higher clinical image quality (Fig 2) and smaller artifacts, which is in accordance with previous studies [6, 8, 26]. This is because increasing the tube voltage increases the effective X-ray energy, which can improve the beam penetration and thus reduce the missing projection data, in turn reducing the artifact areas.

Many groups investigated the usefulness of iterative reconstruction for metal artifact reduction in CT images [27-29]. However, none of the described algorithms was commercially available. In this study, the effect of the commercially available Veo algorithm was investigated. The use of this model based iterative reconstruction algorithm results in significantly improved clinical and physical-technical image quality without compromising low contrast detail (Table 4, Figure 3). The influence of Veo on metal artifacts should be further investigated for other materials and other body regions.

Many studies evaluated the potential of monochromatic CT to reduce image artifacts. Lin et al. [18] found less beam hardening artifacts in monochromatic images compared to the polychromatic images for cranial CT. The study of Zhou et al. [30] indicated the ability of simulated monochromatic images to reduce metal artifact of CT images in patients with implanted metal orthopedic devices after fractures and indicated 130 keV to be the optimal photon energy setting. This positive influence of monochromatic images was also seen by Bamberg et al. [31], which selected 105 keV to be the optimal photon energy setting.

These findings are in general agreement with our findings, where higher keV levels resulted in a better image quality and a reduction of the metal artifacts (Fig 4,7,8). High keV imaging however decreases the low contrast resolution (Fig 5,6).

Another possibility to reduce metal artifacts is the use of dedicated metal artifact reduction software. Wang et al [33] evaluated the dual energy metal artifact reduction tool MARs (GE, Medical Systems) for CT scans of total hip prostheses. The combination of monochromatic reconstruction and MARs diminished the metal artifacts and improved the quality of the CT. However, it was noted that in some circumstances, MARs images were not always superior to monochromatic images without MARs. The results of Brook et al. [34] showed that MARs reconstruction was better than standard reconstruction in nearly 60% of the scanned patients with fiducial seeds implanted for radiation therapy. MARs effectively eliminated blooming of

the fiducial seed, but in some patients introduced far-field artifacts. Wang et al. [14] investigated the use of monochromatic images combined with MARs for the evaluation of pedicle screws in patients with scoliosis and showed that MARs reconstructed images led to suboptimal image quality of pedicle screws.

Overall, the literature supports the value of MARs, though encourages to use MARs with care and suggests MARs reconstructed images to be used as complementary to those with standard reconstruction. In this study, MARs showed to reduce the amount of metal artifacts. However, clinical image quality was clearly negatively affected by the use of MARs resulting in lower IQS values (Fig 4). Clinical images showed an over smoothing of the image resulting in loss of essential image information although less streaking artifacts are visible around the metal insertions in the MARs reconstructed image (Fig 9). Especially in the tongue base and in the proximity of the metal inserts, secondary artifacts were induced. Contrast detail analysis also showed a deterioration of both IQF_{inv} and low contrast visibility in MARs reconstructed images compared to standard reconstructed images (Fig 5,6). Consequently, MARs reconstruction in CT images of the oral cavity to reduce dental hardware metallic artifacts is not sufficient, and may even adversely influence the image quality.

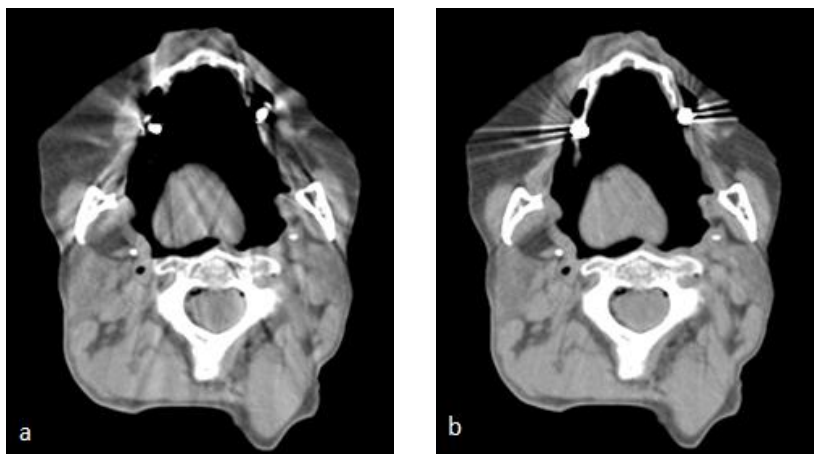


Fig 9: Thiel head CT scan reconstructed with (a) and without (b) MARs. The loss of information around the metal insertions and the deteriorated resolution in MARs reconstructed images is demonstrated although less streaking artifacts are visible around the metal insertions in the MARs reconstructed image. MARs = metal artifact reduction software

When comparing the different investigated techniques, 140 kVp FBP, 120 kVp veo and 140 keV DECT without MARs seems to be promising techniques to reduce metal artifacts caused by dental hardware. In order to determine advantages and disadvantages for all approaches,

several issues must be considered. The radiation dose in the present study was kept constant for all protocols, resulting in no argument for one or the other approach. At a lower energy level, monochromatic images are more suitable for soft tissue detail (Fig 5,6). However, low energy monochromatic images are less efficient in reducing beam hardening artifacts [37]. Limited access to DECT scanners is a limiting factor regarding a widespread application of monochromatic CT image reconstruction. Both, mono-energetic reconstructions and iterative algorithms are easy-to-apply post-processing steps. Nevertheless, both approaches require additional time and software packages. Especially model-based iterative reconstruction results in elevated reconstruction times. Increasing the kVp is possible on every CT device, however, altering the kVp from 120 kVp to 140 kVp hardly has any significant influence (Table 4).

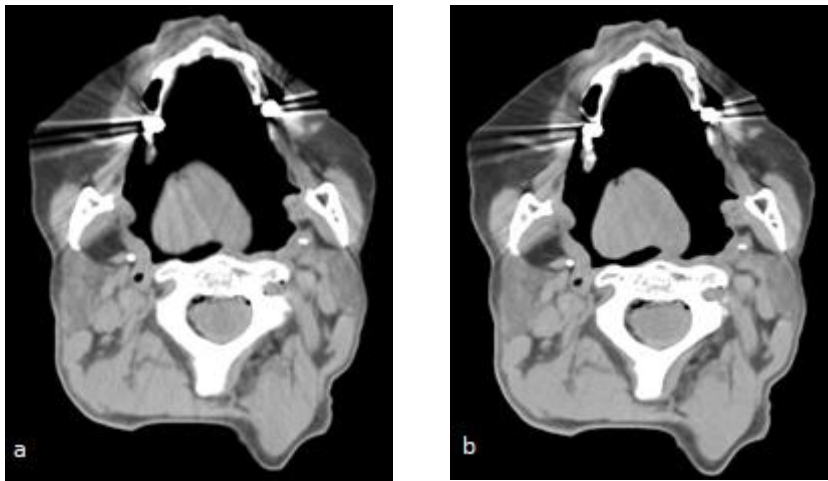
There were some limitations to the present study. Firstly, only cylindrical inserts of 4 mm were studied. Other diameters could result in other efficiency of the investigated metal artifact reduction tools although similar overall trends can be expected. Secondly, only the three most commonly applied materials used as dental implants were investigated, it would be interesting to investigate the influence of different dental restoration materials such as amalgam and gold fillings, which have a higher density compared to the materials used in this study. Fourthly, only a limited amount of artifacts were present in the images since only two metal inserts were used. However, in clinical situations, more severe artifacts can be present. Fifthly, for the clinical assessment of the image quality, only one Thiel embalmed cadaver was used. Sixthly, the results are only valid for the GE Discovery CT 750 HD, since the amount of metal artifacts is dependent on the implemented CT reconstruction algorithms and the use of a single source dual energy scanner.

In conclusion, of all investigated techniques, Veo shows to be most promising, with a significant improvement of both the clinical and physical-technical image quality without adversely affecting contrast detail. High level monochromatic images in dual energy CT do improve the image quality. However, MARs reconstruction in CT images of the oral cavity to reduce dental hardware metallic artifacts is not sufficient, and may even adversely influence the image quality.

References

1. Geets, X., et al., *Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: Comparison between CT-scan and MRI*. Radiotherapy and Oncology, 2005. **77**(1): p. 25-31.
2. Barrett, J.F. and N. Keat, *Artifacts in CT: Recognition and avoidance*. Radiographics, 2004. **24**(6): p. 1679-1691.
3. Lee, M.J., et al., *Overcoming artifacts from metallic orthopedic implants at high-field-strength MR imaging and multidetector CT*. Radiographics, 2007. **27**(3): p. 791-803.
4. Fiala, T.G.S., R.A. Novelline, and M.J. Yaremchuk, *Comparison of Ct Imaging Artifacts from Craniomaxillofacial Internal-Fixation Devices*. Plastic and Reconstructive Surgery, 1993. **92**(7): p. 1227-1232.
5. Haramati, N., et al., *Ct Scans through Metal Scanning Technique Versus Hardware Composition*. Computerized Medical Imaging and Graphics, 1994. **18**(6): p. 429-434.
6. Moon, S.G., et al., *Metal Artifact Reduction by the Alteration of Technical Factors in Multidetector Computed Tomography: A 3-Dimensional Quantitative Assessment*. Journal of Computer Assisted Tomography, 2008. **32**(4): p. 630-633.
7. Lee, I.S., et al., *A pragmatic protocol for reduction in the metal artifact and radiation dose in multislice computed tomography of the spine: Cadaveric evaluation after cervical pedicle screw placement*. Journal of Computer Assisted Tomography, 2007. **31**(4): p. 635-641.
8. Habets, J., et al., *Artifact reduction strategies for prosthetic heart valve CT imaging*. International Journal of Cardiovascular Imaging, 2012. **28**(8): p. 2099-2108.
9. Stradiotti, P., et al., *Metal-related artifacts in instrumented spine. Techniques for reducing artifacts in CT and MRI: state of the art*. European Spine Journal, 2009. **18**: p. S102-S108.
10. Link, T.M., et al., *CT of metal implants: Reduction of artifacts using an extended CT scale technique*. Journal of Computer Assisted Tomography, 2000. **24**(1): p. 165-172.
11. Funama, Y., et al., *A newly-developed metal artifact reduction algorithm improves the visibility of oral cavity lesions on 320-MDCT volume scans*. Physica Medica-European Journal of Medical Physics, 2015. **31**(1): p. 66-71.
12. Zhang, D., X.H. Li, and B. Liu, *Objective characterization of GE Discovery CT750 HD scanner: Gemstone spectral imaging mode*. Medical Physics, 2011. **38**(3): p. 1178-1188.
13. Lee, Y.H., et al., *Metal artefact reduction in gemstone spectral imaging dual-energy CT with and without metal artefact reduction software*. European Radiology, 2012. **22**(6): p. 1331-1340.
14. Wang, Y., et al., *Metal artifacts reduction using monochromatic images from spectral CT: Evaluation of pedicle screws in patients with scoliosis*. European Journal of Radiology, 2013. **82**(8): p. E360-E366.
15. Lewis, M., K. Reid, and A.P. Toms, *Reducing the effects of metal artefact using high keV monoenergetic reconstruction of dual energy CT (DECT) in hip replacements*. Skeletal Radiology, 2013. **42**(2): p. 275-282.
16. Thiel, W., *Ergänzung für die Konservierung ganze Leichen nach W. Thiel*. Annals of Anatomy, 2002. **184**: p. 267-269.
17. Thiel, W., *Die Konservierung ganzer Leichen in natürlichen Farben*. Annals of Anatomy, 1992. **174**: p. 185-195.
18. Lin, X.Z., et al., *High-Definition CT Gemstone Spectral Imaging of the Brain: Initial Results of Selecting Optimal Monochromatic Image for Beam-Hardening Artifacts and Image Noise Reduction*. Journal of Computer Assisted Tomography, 2011. **35**(2): p. 294-297.

19. van der Schaaf, I., et al., *Minimizing clip artifacts in multi CT angiography of clipped patients*. American Journal of Neuroradiology, 2006. **27**(1): p. 60-66.
20. Thijssen, M., K. Bijkerk, and R. van der Burgth, *Manual Contrast-Detail Phantom CDRAD type 2.0. Project Quality Assurance in Radiology, Department of Radiology, University Hospital Nijmegen, St. Radboud, The Netherlands*. 1998.
21. AAPM report no.39, *Specification and acceptance testing of computed tomography scanners*. AAPM, 1993.
22. Samei, E., et al., *Assessment of display performance for medical imaging systems: Executive summary of AAPM TG18 report*. Medical Physics, 2005. **32**(4): p. 1205-1225.
23. Viner, M., et al., *Liver SULmean at FDG PET/CT: Interreader Agreement and Impact of Placement of Volume of Interest*. Radiology, 2013. **267**(2): p. 596-601.
24. Vogel, L., et al., *Intra-rater agreement of the anorectal exam and classification of injury severity in children with spinal cord injury*. Spinal Cord, 2009. **47**(9): p. 687-691.
25. Maroldi, R., et al., *Computed tomography scanning of supraglottic neoplasms: Its cost effective use in preoperative staging*. Academic Radiology, 1996. **3**: p. S57-S59.
26. Chindasombataroen, J., et al., *Quantitative analysis of metallic artifacts caused by dental metals: comparison of cone-beam and multi-detector row CT scanners*. Oral Radiology, 2011. **27**(2): p. 114-120.
27. Boas, F.E. and D. Fleischmann, *Evaluation of Two Iterative Techniques for Reducing Metal Artifacts in Computed Tomography*. Radiology, 2011. **259**(3): p. 894-902.
28. Kondo, A., et al., *Iterative correction applied to streak artifact reduction in an X-ray computed tomography image of the dento-alveolar region*. Oral Radiology, 2010. **26**(1): p. 61-65.
29. Dong, J., et al., *Metal-induced streak artifact reduction using iterative reconstruction algorithms in x-ray computed tomography image of the dentoalveolar region*. Oral Surgery Oral Medicine Oral Pathology Oral Radiology, 2013. **115**(2): p. E63-E73.
30. Zhou, C.S., et al., *Monoenergetic Imaging of Dual-energy CT Reduces Artifacts from Implanted Metal Orthopedic Devices in Patients with Fractures*. Academic Radiology, 2011. **18**(10): p. 1252-1257.
31. Bamberg, F., et al., *Metal artifact reduction by dual energy computed tomography using monoenergetic extrapolation*. European Radiology, 2011. **21**(7): p. 1424-1429.
32. Yu, L.F., S. Leng, and C.H. McCollough, *Dual-Energy CT-Based Monochromatic Imaging*. American Journal of Roentgenology, 2012. **199**(5): p. S9-S15.
33. Wang, F.D., et al., *Reduction of Metal Artifacts From Alloy Hip Prostheses in Computer Tomography*. Journal of Computer Assisted Tomography, 2014. **38**(6): p. 828-833.
34. Brook, O.R., et al., *Spectral CT with Metal Artifacts Reduction Software for Improvement of Tumor Visibility in the Vicinity of Gold Fiducial Markers*. Radiology, 2012. **263**(3): p. 696-705.
35. Han, S.C., et al., *Metal Artifact Reduction Software Used With Abdominopelvic Dual- Energy CT of Patients With Metal Hip Prostheses: Assessment of Image Quality and Clinical Feasibility*. American Journal of Roentgenology, 2014. **203**(4): p. 788-795.
36. Yu, L.F., et al., *Virtual monochromatic imaging in dual-source dual-energy CT: Radiation dose and image quality*. Medical Physics, 2011. **38**(12): p. 6371-6379.
37. Pessis, E., et al., *Virtual Monochromatic Spectral Imaging with Fast Kilovoltage Switching: Reduction of Metal Artifacts at CT*. Radiographics, 2013. **33**(2): p. 573-583.

Addendum:

Thiel head CT scan at 120 kVp reconstructed with (a) FBP and (b) Veo. Veo decreases the induced metal artifacts resulting in an improvement of all investigated parameters.

References

1. Cowen, A.R., S.M. Kengyelics, and A.G. Davies, *Solid-state, flat-panel, digital radiography detectors and their physical imaging characteristics*. Clinical Radiology, 2008. **63**(5): p. 487-498.
2. Dobbins, J.T., *Handbook of Medical Imaging. Vol 1, Physics and Psychophysics: Image quality metric for digital system*. 2000, Bellingham, Washington: SPIE Press.
3. Hermann, T.L., et al., *Best Practices in Digital Radiography*. American Society of Radiologic Technologists, 2012.
4. Dose Datamed, *DDM2 Project Report Part 1: European Population Dose: Study on European Population Doses from Medical Exposure*. 2014.
5. RIZIV, *Data received by the department of medical care of the RIZIV. Thanks to Engels H*. 2014.
6. Goodenough, D.J., *Handbook of Medical Imaging. Vol 1, Physics and Psychophysics: Tomographic imaging*. 2000, Bellingham, Washington: SPIE Press.
7. Bushberg, J., et al., *The essential physics of medical imaging*. 2nd ed. 2002, Philadelphia, PA: Lippincott Williams & Wilkens.
8. Carlton, R.R. and A.M. Adler, *Principles of Radiographic Imaging, An Art and a Science*. 4th ed. 2006, New York (US): Thomas Delmar Learning.
9. Dougeni, E., K. Faulkner, and G. Panayiotakis, *A review of patient dose and optimisation methods in adult and paediatric CT scanning*. European Journal of Radiology, 2012. **81**(4): p. E665-E683.
10. American Association of Physicists in Medicine. *The measurement, reporting and management of radiation dose in CT*. AAPM Report Task Group 23, Report no 96, 2008.
11. Singh, S., et al., *Radiation Dose Optimization and Thoracic Computed Tomography*. Radiologic Clinics of North America, 2014. **52**(1): p. 1-+.
12. Winklehner, A., et al., *Raw data-based iterative reconstruction in body CTA: evaluation of radiation dose saving potential*. European Radiology, 2011. **21**(12): p. 2521-2526.
13. Lee, T.-Y. and R.K. Chhem, *Impact of new technologies on dose reduction in CT*. European Journal of Radiology, 2010. **76**: p. 28-35.
14. Beister, M., D. Kolditz, and W.A. Kalender, *Iterative reconstruction methods in X-ray CT*. Physica Medica-European Journal of Medical Physics, 2012. **28**(2): p. 94-108.
15. Baker, M.E., et al., *Contrast-to-Noise Ratio and Low-Contrast Object Resolution on Full- and Low-Dose MDCT: SAFIRE Versus Filtered Back Projection in a Low-Contrast Object Phantom and in the Liver*. American Journal of Roentgenology, 2012. **199**(1): p. 8-18.
16. Prakash, P., et al., *Diffuse Lung Disease: CT of the Chest with Adaptive Statistical Iterative Reconstruction Technique*. Radiology, 2010. **256**(1): p. 261-269.
17. Ghetti, C., O. Ortenzia, and G. Serrelli, *CT iterative reconstruction in image space: A phantom study*. Physica Medica-European Journal of Medical Physics, 2012. **28**(2): p. 161-165.
18. Khawaja, R.D.A., et al., *Computed Tomography (CT) of the Chest at Less Than 1 mSv: An Ongoing Prospective Clinical Trial of Chest CT at Submillisievert Radiation Doses with Iterative Model Image Reconstruction and iDose(4) Technique*. Journal of Computer Assisted Tomography, 2014. **38**(4): p. 613-619.
19. Nishio, M., et al., *Emphysema quantification on low-dose CT using percentage of low-attenuation volume and size distribution of low-attenuation lung regions: Effects of adaptive*

- iterative dose reduction using 3D processing*. European Journal of Radiology, 2014. **83**(12): p. 2268-2276.
20. Mievile, F.A., et al., *Iterative reconstruction methods in two different MDCT scanners: Physical metrics and 4-alternative forced-choice detectability experiments - A phantom approach*. Physica Medica-European Journal of Medical Physics, 2013. **29**(1): p. 99-110.
 21. Kulkarni, N.M., et al., *Emerging Technologies in CT-Radiation Dose Reduction and Dual-Energy CT*. Seminars in Roentgenology, 2013. **48**(3): p. 192-202.
 22. Johnson, T.R.C., *Dual-Energy CT: General Principles*. American Journal of Roentgenology, 2012. **199**(5): p. S3-S8.
 23. Zhang, L.J., et al., *Dual-Energy CT Angiography in the Evaluation of Intracranial Aneurysms: Image Quality, Radiation Dose, and Comparison With 3D Rotational Digital Subtraction Angiography*. American Journal of Roentgenology, 2010. **194**(1): p. 23-30.
 24. Watanabe, Y., et al., *Dual-energy direct bone removal CT angiography for evaluation of intracranial aneurysm or stenosis: comparison with conventional digital subtraction angiography*. European Radiology, 2009. **19**(4): p. 1019-1024.
 25. Brown, C.L., et al., *Dual-energy CT iodine overlay technique for characterization of renal masses as cyst or solid: a phantom feasibility study*. European Radiology, 2009. **19**(5): p. 1289-1295.
 26. Koonce, J.D., et al., *Accuracy of dual-energy computed tomography for the measurement of iodine concentration using cardiac CT protocols: validation in a phantom model*. European Radiology, 2014. **24**(2): p. 512-518.
 27. Karlo, C.A., et al., *Split-bolus dual-energy CT urography: protocol optimization and diagnostic performance for the detection of urinary stones*. Abdominal Imaging, 2013. **38**(5): p. 1136-1143.
 28. Zhang, L.J., et al., *Liver virtual non-enhanced CT with dual-source, dual-energy CT: a preliminary study*. European Radiology, 2010. **20**(9): p. 2257-2264.
 29. Ferda, J., et al., *The assessment of intracranial bleeding with virtual unenhanced imaging by means of dual-energy CT angiography*. European Radiology, 2009. **19**(10): p. 2518-2522.
 30. Bamberg, F., et al., *Metal artifact reduction by dual energy computed tomography using monoenergetic extrapolation*. European Radiology, 2011. **21**(7): p. 1424-1429.
 31. Lin, X.Z., et al., *High-Definition CT Gemstone Spectral Imaging of the Brain: Initial Results of Selecting Optimal Monochromatic Image for Beam-Hardening Artifacts and Image Noise Reduction*. Journal of Computer Assisted Tomography, 2011. **35**(2): p. 294-297.
 32. Zhou, C.S., et al., *Monoenergetic Imaging of Dual-energy CT Reduces Artifacts from Implanted Metal Orthopedic Devices in Patients with Fractures*. Academic Radiology, 2011. **18**(10): p. 1252-1257.
 33. Tawfik, A.M., et al., *Image Quality and Radiation Dose of Dual-Energy CT of the Head and Neck Compared with a Standard 120-kVp Acquisition*. American Journal of Neuroradiology, 2011. **32**(11): p. 1994-1999.
 34. Fung, G.S.K., et al., *Differentiation of Kidney Stones Using Dual-Energy CT With and Without a Tin Filter*. American Journal of Roentgenology, 2012. **198**(6): p. 1380-1386.
 35. Kaufmann, S., et al., *Tin-filter Enhanced Dual-Energy-CT: Image Quality and Accuracy of CT Numbers in Virtual Noncontrast Imaging*. Academic Radiology, 2013. **20**(5): p. 596-603.
 36. Barrett, J.F. and N. Keat, *Artifacts in CT: Recognition and avoidance*. Radiographics, 2004. **24**(6): p. 1679-1691.

Part IV:

General discussion

8

Strength and weakness of different image quality parameters

There is a large variety in parameters describing the image quality of medical X-ray images. However, depending on the intended purpose, the correct image quality parameter has to be selected, taking into account their different strengths and weaknesses.

8.1 Physical-technical image quality parameters

8.1.1 Noise

Noise and noise related properties of a radiological image are of great importance to diagnostic performance. Noise is based on the standard deviation of pixel values in an image, while NPS is a metric of image quality used to measure the noise characteristics and patterns in all frequencies of the image. In this way NPS provides us with a more complete description of noise in an image. Noise and noise related parameters are used for many applications, such as benchmarking image quality across systems, optimizing acquisition and reconstruction parameters to improve dose efficiency, and predicting observer performance. The question is of course if the properties of noise and NPS are suitable to draw such type of conclusions. The detection of lesions or abnormalities in medical images is impaired greatly by the presence of image noise. One component of medical image noise is quantum noise, the variability in the image's intensity distribution due to the statistical fluctuations in the number of photons

reaching the image receptor [1]. A second component of medical image noise is anatomical noise, which originates from anatomic structures in the image that are irrelevant to the detection task [1, 2]. Noise and noise related properties are quantified on structure-free images which implies that the radiological task of detection of lesions or abnormalities is quantum noise limited rather than limited by the projected anatomy. This assumption is obviously not true for many common tasks in diagnostic radiology [1, 3-5]. Furthermore, when the anatomical noise dominates, the image quality depends less on the dose than expected from quantum noise considerations [5].

Another drawback of using noise as an image quality parameter is that noise is commonly estimated over a ROI in a single image, employing assumptions of linearity and stationarity [6]. However, in newly developed reconstruction methods as iterative reconstruction (IR), the system is potentially nonlinear which implicates that noise is highly spatially dependent and that image resolution depends on contrast [7, 8]. So the evaluation of noise and resolution depends on the properties of the imaging task. As a result, task-independent metrics such as noise or NPS are no longer adequate for evaluations of IR image quality [9, 10].

These findings are supported by our research. In chapter 6, the results of an optimization study for IR, based on different image quality parameters is presented. Potential dose reduction based on noise, is explicitly higher than the potential dose reduction based on clinical image quality assessment. Since noise was evaluated in homogeneous phantom images, the assumption was made that the radiological task of scoring normal anatomical structures in a chest CT is limited by the quantum noise. Moreover, using noise to evaluate the iterative reconstruction method SAFIRE assumes a linearity of the algorithm which is not the case [9, 11, 12]. Our results demonstrate that for IR algorithms, noise obtained in uniform phantom images, clearly is not appropriate to quantify the potential dose reduction.

In literature, IR is mostly assessed with task independent metrics, such as voxel noise and contrast-to-noise ratio [10, 13, 14]. As a result, usually a single dose reduction potential independent of tasks is prescribed to IR. Pickhardt et al. [15] investigated the use of model based iterative reconstruction (MBIR) for abdominal CT based on noise measurements and lesion detection. MBIR resulted in lower image noise but depicted fewer lesions than FBP images. Chen et al. [9] demonstrated that tasks with lower contrast and smaller size have higher absolute dose reduction potential. Both articles highlight that noise based measurements are not sufficient to optimize nonlinear IR based algorithms. This stresses the

importance of evaluating image quality in IR with respect to the task. To accommodate IR's nonlinearity, Chen et al. [12] extended the conventional concepts of NPS to magnitude-dependent NPS_{task} , which characterizes the noise properties at a specific noise magnitude. The measurement of NPS_{task} enables the comparison of IR performance in terms of noise.

Therefore, noise measurements in a structure-free image or for nonlinear systems are not valid for clinical image optimization processes. However, for comparison between different reconstruction techniques, noise measurements can be used as an objective image quality parameter [16]. Similarly, noise related factors such as NPS can be of interest for describing the physical-technical performance of radiological imaging equipment for acceptance and consistency testing.

8.1.2 Resolution

The evaluation of spatial resolution of imaging systems plays a central role in imaging performance evaluation [17]. One of the most comprehensive metrics used to measure and report spatial resolution of imaging systems is the MTF. The MTF provides a measure of how well the system transfers contrast across spatial-frequencies. The MTF is also essential to evaluating other key imaging performance metrics such as the DQE. The MTF is a useful quantity for linear systems, where the imaging system's response to an arbitrary object can be determined by convolving the true object and the point spread function [6]. This is not true for nonlinear IR algorithms. As the MTF is not well defined for images reconstructed by IR, it is of limited utility in assessing the quality of these images. The introduction of iterative reconstruction systems has posed a challenge in our ability to assess spatial resolution.

Current MTF assessment methods only focus on linear systems with a single contrast and noise level. As a result, Richard et al. [7] presented a method for measuring the MTF_{task} for CT systems for different reconstruction algorithms and across a range of noise and contrast levels. The MTF was applied not as a generic reflection of system resolution but rather as an object specific quantity. This was done so to accommodate the nonlinearity of iterative reconstruction methods.

8.1.3 Contrast

The CNR is a useful metric for describing the signal amplitude relative to the ambient noise for simple and largely homogeneous objects. However, the CNR depends only on contrast and noise. Actual signal detectability also depends on factors including signal size, shape and density distribution; background level, variability and correlation; the variance and covariance of measurement noise; spatial resolutions; and the observer and detection strategy used. The CNR can be useful in some simple situations, e.g. determining thresholds of contrast agents at which signals on a test phantom become visible. However, the CNR is in general not a complete description of an observer's ability to detect lesions, and this is even more true for IR images which are more likely to be nonlinear and nonstationary [6].

The results of the study described in chapter 6 also demonstrate the limited value of CNR to describe the entire image quality assessment. Potential dose reductions based on CNR are explicitly higher than the potential dose reduction based on clinical image quality assessment. The inability of CNR to give a complete description of an observer's ability to detect lesions is also shown by Schindera et al. [18]. In this study, low contrast detectability was assessed by means of simulated lesions in a liver phantom. Moreover, objective image quality was assessed by calculating CNR values. The results of the phantom study indicated that use of IR does not improve the low contrast detectability of simulated liver lesions. However, CNR values were the same in 100% dose FBP images and in 20% IR images. They concluded that IR substantially improves the quantitative image quality based on CNR measurements, but does not improve the diagnostic effectiveness. Consequently, optimization procedures based solely on CNR is inappropriate since changes in IR are not necessarily reflected in clinical low contrast detectability. Moreover, the image quality may still deteriorate in terms of spatial resolution and artifacts [10].

8.1.4 Contrast-detail

Previous mentioned image quality parameters e.g. noise, resolution and contrast, are objective physical parameters for assessment of IQ based solely on measurements of image data sets. However, since these parameters do not include the effect of a human observer, they cannot be used to describe the entire imaging chain. Due to its crucial role in the medical diagnosis process, human decision criteria are a fundamental element necessary to include in

the imaging chain when assessing IQ. An alternative for the assessment of physical-technical image quality is the use of contrast-detail phantoms. Evaluation of these phantoms includes a human observer and provides a quantitative measure in terms of two fundamental parameters: low contrast and small detail detectability. The utilization of CD phantoms is heavily reliant on 'the Rose model', as put forward by Albert Rose in 1948 [19]. Rose stated that there is an inverse relationship between the linear size of an object and the minimum threshold contrast required for its detection. The use of the Rose model in radiology implies that the tasks are quantum-noise limited rather than limited by the anatomical noise. As described in paragraph 8.1.1, this assumption is obviously not true for many common tasks in diagnostic radiology. Contrast detail phantoms do not contain any anatomical structures (either normal anatomical background nor pathology) and thus do not resemble the clinical task of the observer. Consequently, contrast detail studies cannot be applied for optimization purposes. However, various authors reported its extended use in optimization studies [20, 21]. Nevertheless, contrast detail studies provide useful information on equipment performance [22]. The use of contrast-detail images is a practical approach primarily implemented for routine quality control and constancy checks. The current fourth edition of the European Guidelines for Quality Assurance in Mammography Screening [23] specifies minimum performance in terms of image quality and radiation dose. The CDMAM contrast detail phantom is routinely used in European quality control programs. However, the use of contrast detail phantoms suffers from several disadvantages. Subjective image quality evaluation methods based on visual interpretation depend on the variable and subjective nature of human observer decisions. Previous studies have revealed inter- and intra-observer variability in scoring of contrast-detail images [24, 25]. This variability may affect the result of the threshold visibility test and thus limits the ability of accurately assessing system performance. These limitations can be partially overcome if several observers are used. Nevertheless, the interpretation by multiple observers is impractical for routine quality assurance. Moreover, scoring of contrast-detail images is very time consuming. Therefore, software tools which automatically evaluate contrast-detail images have been developed to suppress inter-observer errors [26, 27].

Pascoal et al. [28] investigated the CDRAD analyzer for automated quality assessment of contrast detail images compared with subjective visual assessment. The software proved to be more sensitive than the average observer in the detection of the lower contrast details. Furthermore, in this study, the use of a single image quality figure such as the IQF_{inv} , is

discouraged for absolute comparison of image quality, because important variations in detection performance with varying detail diameter and hole depths cannot be perceived. They may however be considered as useful indicators of overall image quality for relative comparisons as they provide a quick method of exploring trends of image quality performance.

The results of the study described in chapter 5 demonstrate the correlation between automated and manual readings of the CDRAD phantom. A mean increase of 19% was seen in the measurements of the automatically obtained IQF_{inv} values compared with the value derived by the human observers. Therefore, the software tool was more sensitive than the human observers in the correct detection of objects in the contrast-detail phantom. However, as reported by Pascoal et al. [28], changes in the significance level of the software tool, defined by the user prior to analysis can result in higher or lower IQF_{inv} values.

Contrast detail analysis was used in all studies described in Part III. In chapter 5, an excellent correlation was found between IQF_{inv} values and clinical imaging quality by means of an VGA score. This study supports the value of contrast detail phantom analysis for evaluating clinical quality in chest radiography. In chapter 6, a contrast detail study was performed to assess the performance of iterative reconstruction techniques in CT. Among different physical parameters, contrast detail seems the best indicator for clinical performance. In chapter 7, metal artifact reduction tools were investigated by means of noise related parameters, contrast-detail and clinical image quality. Contrast detail results appeared to indicate the clinical effect of the different scanning and reconstruction settings better compared to the noise related image quality parameters.

8.2 ROC and VGAS

An alternative to the previous described parameters is task-based image quality assessment. Both VGA and ROC or ROC related studies are based on the evaluation of certain tasks. ROC and ROC related studies are currently considered one of the best methods to quantify and report diagnostic performance. They measure the ability of an observer to detect and correctly interpret pathological structures. However, there are several important disadvantages to ROC studies [29]. Firstly, large number of patient images and pathological lesions are required to obtain a sufficient statistical power. Secondly, the true health state of the patient must be

known to classify an image as normal or abnormal. Thirdly, the result of a ROC study only applies to a certain kind of lesion which had to be detected. To be representative for clinical scenarios, task-based assessments should be performed for a variety of tasks, resulting in a large number of images and image evaluations. Fourthly, the time consumption for the readers, but also for the preparation of sets of images with subtle lesions is extensive [30]. These requirements make the measurement of clinical performance by means of ROC studies difficult in practice and very time consuming.

A simpler method has been published by the European Commission [31] and is based upon the scoring of image quality by inspecting normal anatomical structures. Good visibility of these normal structures is considered to define appropriate image quality and accurate diagnosis. The set of anatomical criteria is specific for the given type of examination and VGA is typically applied to evaluate them. An advantage of VGA methods is that contrary to the ROC related methods, it is the normal anatomy present in almost every patient image which is used for the evaluation of image quality. This property of VGA can also be considered as a disadvantage, since the diagnostic task is missing [29]. However, it seems clear that the visibility of normal anatomy is of great diagnostic significance to radiologists. VGA experiments provide an overall assessment of visibility of normal structures in terms of contrast, noise and sharpness. Demonstration of expected structures and patterns is reassuring to the interpreting physician and allows pathology to be excluded more reliably [32]. For example, a processing algorithm with a high noise perception by the radiologists will most likely not be widely adopted in clinical practice, even if good diagnostic accuracy had been proved [30]. VGA studies are relatively easy to conduct, especially in comparison with ROC studies, and the time consumption is moderate, at least for the observers. It is realistic to believe that this method can be implemented at almost any hospital [33].

Correlation between VGA experiments and ROC studies has previously been investigated, but the results are contradictory [29, 34]. In 2000, Tingberg et al. [34] investigated the diagnostic quality of radiographs of the lumbar spine with both VGA and FFE methods. Lumbar spine images were manipulated by adding artificial lesions, and image processing was applied to simulate image appearance from three different screen film combinations. The ranking of the images was the same when using ROC or VGA. A linear correlation was found between the results of the two methods. The uncertainties in the VGAS values were lower than the uncertainties in the FFE values, indicating VGA to be a robust method for evaluation of the image quality. In 2005, the same group performed a second study on the same subject [29]. In

this paper the authors reported that images with greater contrast than the original had significantly improved image quality in the VGA study although this did not affect detectability in the FFE experiment. From this, it is doubtful that the fulfillment of the investigated quality criteria for the visualization of normal structures in a VGA experiment is sufficient for accurate detection of pathology. The use of VGA is motivated by the assumption that the level of visibility of anatomical and pathological structures are connected, so that if the visibility of the normal anatomy is increased, the visibility of pathological structures is also increased. In the described study, increased noise in the images was easily detected by the radiologists in the VGA study, but did not affect the detection of the lesions in the FFE study. Good visibility of normal anatomy seems to be a needed, but not a sufficient, requirement for lesion detection. Basically the essential difference between both methods can be found in the following attitudes: or the radiologist feels confident with the global appearance of an image or it should be tested whether a new exposure or reconstruction setting also guarantees better lesion detection [30]. Both aspects are very relevant for the daily use of an imaging system and therefore both testing methods, VGA and ROC are useful. Correlations between FFE and VGA can probably only be found in a few special cases where the pathological structures for investigation in the FFE study are of the same size, shape, texture and contrast as the normal anatomy used in the VGA study. If the pathological structures that are searched for in the ROC study are different with respect to size, shape texture and/or contrast, to the anatomical structures used in the VGA study it cannot be taken for granted that correlation would exist. Since the methodology and the results of the two methods are quite different it is necessary to understand their advantages and disadvantages to be able to select the method that is most appropriate in a given situation.

8.3 Model observers

The evaluation of a new reconstruction algorithm or processing algorithms is frequently assessed with task-independent metrics such as noise, CNR, MTF and contrast-detail. As a result, a single dose reduction potential independent of tasks is obtained. However, a method of validating dose reduction claims must take into account clinical tasks. To be representative of clinical scenarios, task-based assessments should be performed under a variety of conditions and for a variety of lesions. As a result, these studies can result in a very large number of images and image evaluations, making model observers a valuable tool for image

quality assessment. While human observers are certainly a possibility, in addition to being more cost-effective, model observers are immune to effects like reader learning, fatigue, and inter- and intra-observer variability that can be issues for humans [6]. Model observers can also contribute to an increased reliability of physical-technical measurements such as contrast-detail images. There is a long history in the development of models to predict human performance for visual detection of signal embedded in white noise backgrounds [35]. More recent work has been concentrating on applying a variety of model observers to more complex computer-generated backgrounds including lumpy backgrounds [36, 37]. However, it is important to predict human performance detecting signals in real backgrounds such as detection of a tumor in medical images. Accomplishing this goal requires extending the application of model observers from synthetic backgrounds to real anatomic backgrounds [38, 39]. One of the difficulties in extending model observers to real backgrounds is the lack of statistical information of the images. Different models are developed which differ in the amount of information they use. For example, some models use information about the signal; others use information about the signal and the background statistics (*Figure 8-1*). Models also differ in the components they include to reflect constraints imposed by the human visual system. A fundamental component that distinguishes two groups of noise models is processing by a set of channels that are spatial-frequency tuned. These channels are intended to reflect the response properties of receptive fields of cells in the visual cortex which selectively respond to signals with a given spatial frequency. Therefore, models are divided in two categories: non-channelized models like the non-prewhitening matched filter observer with an eye-filter (NPWE) [6, 40-43] and the Hotelling observer [40, 44] and channelized models like the channelized Hotelling observer [6, 40, 42, 44].

It is important to realize, that optimization studies based on dose reduction potential can lead to different results for different model observer studies [9]. Results are always dependent on the type of investigated lesion and the selected background and observer model. Some studies reflect the potential related to clinical performance, while others represent the maximum performance that can be achieved in ideal situations. Moreover, for every imaging modality and every observer model, the correlation between model observer and human observer has to be established [45, 46]. Consequently, model observers are definitely an interesting tool for image quality analysis and optimization studies. However it is important to be aware that every step (selection of signal, background and observer model) influences the outcome of the study.

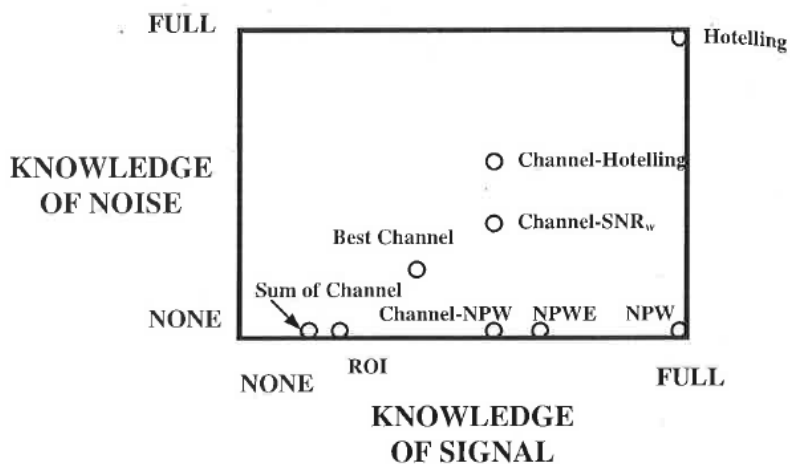


Figure 8-1: A quantitative plot of model observers based on the knowledge they use about the signal and background statistics to build the template. Particular signals might favor one model over another. The diagram ranks the models on their use of knowledge to derive templates averaged over all possible signals. [47]

The value of Thiel embalmed human cadavers

To analyze both physical-technical and clinical image quality, a variety of phantoms are available. Phantoms to evaluate clinical image quality performance are partly described in paragraph 2.3. Phantoms to describe the physical technical image quality are usually made out of a uniform material. However, as described before (see paragraph 8.1.1), noise is the set of all harmful signals superimposed on a useful signal. The useful signal encodes the diagnostic information, whereas the noise represents a disturbance in understanding the information transmitted by the useful signal. For a medical imaging system, this definition applies to both random fluctuations associated with image detector and display as well as those associated with anatomical variations. One of the difficulties encountered with anatomical noise arises from the fact that it depends on the diagnostic task [5]. It can only be defined in association with the signal of interest. For instance, when examining blood vessels on chest images, the ribs can be considered as a form of anatomical noise. When examining small rib fractures, the blood vessels become anatomical noise and the ribs contain the signal of interest [48]. This stresses the importance of the presence of anatomical structures in the phantom for performing optimization studies.

Since a lot of radiological examinations are limited by the anatomic noise rather than the quantum noise, optimization studies need to be performed in images containing realistic anatomical features and specific diagnostic tasks. Clinical image quality optimization based on ROC or model observer studies currently often make use of anthropomorphic phantoms [49, 50]. Although these phantoms represent better the human anatomy compared to uniform phantoms used for physical-technical measurements, they still have some important

drawbacks. Phantoms which are divided into sections cut orthogonal to the cranial–caudal axis of the phantom inevitably contain air gaps between the different slices which can significantly affect the image quality of CT images, particularly in the areas close to these gaps [51]. Furthermore, anthropomorphic phantoms are rather limited in their composition being composed of just a small fraction of tissues, organs and densities present in the real patient. Another possibility for ROC or model observer studies is the superimposition of lesions on real patient images [52]. However, this implies that the actual truth state of the patient has to be known. Another possibility is the use of embalmed cadavers. However, formalin embalmed cadavers suffer from profound changes of density of organs and tissues. Consequently, X-ray images of formalin embalmed cadavers show significant differences with patient images. Thiel embalmed human cadavers can offer an interesting alternative. Thiel-fixed specimens have outstandingly lifelike visual properties without having the irritating odor associated with formalin embalming procedures. These embalmed cadavers have, just like real patients, the benefit of a realistic anatomical background. Moreover, Thiel embalmed cadavers have important advantages over real patients: knowing the actual truth state of these cadavers is not a problem and these cadavers can be exposed to multiple acquisitions without ethical problems from radiation protection point of view.

Thiel embalmed human cadavers were used in all articles described in Part III of this thesis. The similarity of both planar radiography images (*Figure 9-1*) and tomographic images (*Figure 9-2*) of Thiel embalmed human cadavers with corresponding images of a real patient was assessed by different radiologists. This comparison was made for the chest region and the head and neck region. For both acquisition methods and both body regions, Thiel images were considered appropriate to assess the clinical image quality. Consequently, Thiel embalmed human cadavers can be a valuable tool for optimization studies based on ROC or model observers.

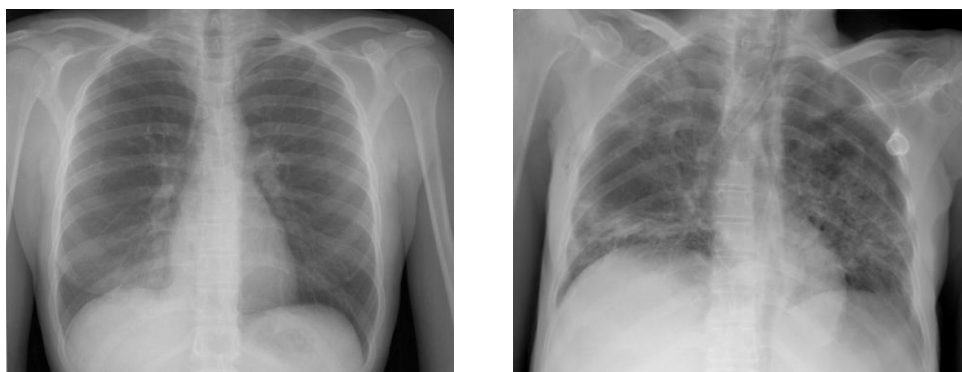


Figure 9-1: Chest radiograph of a patient (left) and a Thiel embalmed cadaver (right).

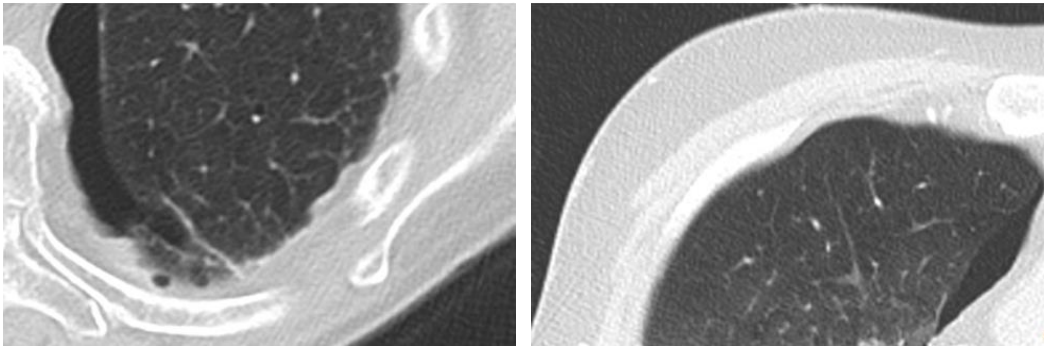


Figure 9-2: Normal lung parenchyma in both a patient (left) and a Thiel cadaver (right) chest CT image.

Application in medical X-ray imaging is not the only scope for Thiel embalmed cadavers. Human cadavers have become an important resource for surgical training. Training on cadaver specimens gives the opportunity to perform major laparoscopic operations while respecting human ethics. Difficult endoscopic and surgical procedures can be trained step by step under real anatomical situations. Therefore, the use of Thiel embalmed cadavers is probably the best method for surgical trainees to learn basic laparoscopic surgery and for advanced surgeons to improve their performance [53]. Laparoscopic procedures are not the only surgical procedures where Thiel embalmed cadavers showed to have a significant value. The significance of Thiel embalmed cadavers has been reported in several areas such as oral surgery [54], micro vascular exercise [55], ultrasound guided regional anesthesia [56] and several urology based procedures [57]. Thiel embalmed cadavers are promising as a model for revascularization studies. Chevallier et al. performed total body postmortem circulation using a heart-lung machine [58]. For this thesis, revascularization of 2 Thiel embalmed cadavers was performed to investigate the possibility of performing abdomen contrast CT scans. First results seemed promising, however, further optimization is necessary. Thiel fixed specimens are not suitable for histological investigations or biomechanical testing [59].

Furthermore, the use of Thiel cadavers complies with the new challenges of shorter training, limited student exposure to anatomy in medical school curricula, and patient safety [60]. Embalming helps to enhance the durability of the cadavers, though at the cost of altering certain tissue characteristics. Formalin fixed cadavers are of little importance in teaching laparoscopic surgery, as the consistency and texture of tissues are significantly altered [53]. Training on Thiel cadavers, compared to animal models, offers an anatomy identical to that found in patients. Moreover, a same cadaver can be used for several different procedures.

However, the use of Thiel embalmed cadavers also has some disadvantages: the Thiel embalming method is expensive and requires an elaborate setup.

Clinical versus physical image quality assessment

There are many tasks in radiology departments which involve assessment of image quality. Equipment purchasing is partly based on performance specifications, acceptance testing verifies that the system fulfills the specified performance criteria, constancy testing focuses on noticing any changes in the imaging system, clinical testing concentrates on the fulfillment of clinical needs, and optimization processes attempt to find best ways to use the imaging system for clinical purposes [32]. These different tasks are best performed by different assessment methods and the outcome is often referred to as physical-technical image quality or clinical image quality, according to the method used.

Although establishing the link between physical-technical image quality measures and clinical utility has been pursued for decades, the relationship between the results of physical-technical measurements and clinical performance is not yet fully established and understood. The relationship between diagnostic performance and physical image quality has been discussed in the International Commission on Radiation Units and Measurements (ICRU), where it was described by a simplified graph (*Figure 10-1*) [61]. The graph shows that if the level of physical image quality is extremely low, the image can provide no information for the diagnosis. When the physical image quality improves, important radiological patterns become recognizable and diagnostic performance improves. Beyond a certain level of physical image quality, the performance will saturate since all important features are already visible [32]. This figure may be representative for a very general approach for the correlation between physical-technical and clinical image quality. However it does not contribute to the essential question: can optimization studies based on physical-technical image quality parameters contribute to the

clinical optimization process? It is therefore important to establish the relation between the physical measures of image quality and clinical image quality as evaluated by radiologists.

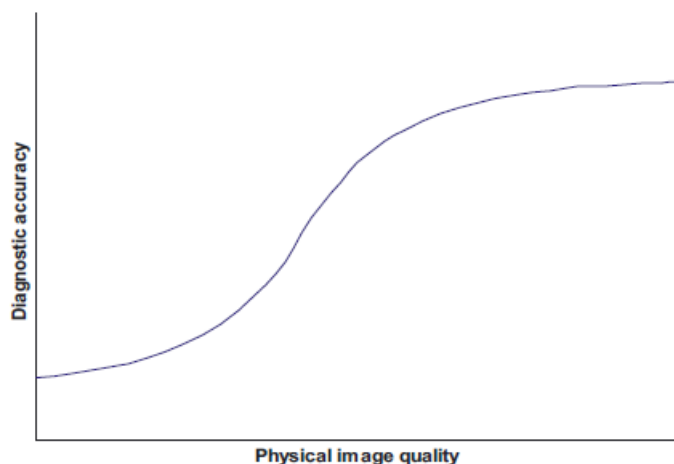


Figure 10-1: Simplified qualitative relationship between physical image quality and diagnostic performance [32]

Although the importance of correlating physical-technical and clinical image quality is highly recognized [32], only a few studies really tried to establish a link between the two. Sandborg et al. [62] studied this correlation in chest and lumbar spine screen-film radiography based on contrast and SNR of specified details and a VGA. Their results show significant correlation of blood vessel contrast and the clinical evaluations in chest imaging. The correlation of the SNR of the blood vessel and clinical evaluation was lower. In lumbar spine imaging, both contrast and SNR of small soft tissue cavities in bone seemed to be good indicators for the clinical image quality. However, as screen-film radiography is no longer applied, these conclusions are no longer relevant and this kind of studies should be repeated for digital radiography.

In a subsequent study of Sandborg et al. [63], the effect of X-ray tube voltage on digital chest and pelvis radiography was investigated. Clinical image quality was evaluated on an anthropomorphic phantom using the VGA method and physical image quality was described by the SNR for a number of small details at various locations in the phantom. Both the clinical and the physical evaluation resulted in a monotone decrease of image quality with increasing X-ray tube voltage. They found a positive linear relationship between the results of the two evaluation methods (chest PA: $r^2 = 0.91$, pelvis: $r^2 = 0.94$), indicating that the SNR is strongly related to the radiologists' grading of the images.

Redlich et al. [64] assessed several chest radiography systems by measuring the DQE and performing a VGA and a ROC study. It was noted that the ranking of the image quality of the systems was nearly the same with all these assessment systems. However, no real correlation was calculated between the different approaches.

In chapter 5 of this thesis, the correlation between the results of a contrast-detail analysis and clinical image quality in chest radiography was investigated. This study showed an excellent correlation between IQF_{inv} and VGAS by means of a spearman correlation coefficient of $r = 0.916$ ($p < 0.001$). This highlights the relevance of contrast-detail analysis as an important physical-technical image quality assessment tool. Because the correlation between VGAS and ROC is questionable, the excellent correlation between contrast-detail and VGAS does not automatically mean that contrast-detail analysis is an appropriate tool for lesion detection optimization studies. However, it is highly relevant for constancy checks and to indicate improvement or deterioration of the normal anatomy which is of great diagnostic significance for the radiologist (see paragraph 8.2).

In chapter 6, a dose optimization study for the use of IR in CT is presented. Potential dose reductions were calculated for different physical-technical and clinical image quality parameters. However, the actual correlation between physical-technical and clinical image quality was not quantified by a correlation coefficient. This study shows clearly that noise and noise related parameters are difficult to link with the clinical image quality. Once again, the contrast-detail analysis seemed most appropriate to predict the outcome of clinical image quality studies based on the normal anatomy.

Finding a correlation between clinical and physical image quality is extremely difficult. Therefore, it is very important to understand the difference between these different approaches. In this thesis, an overview has been made of some common methods for evaluating imaging systems to discuss the validity and reliability associated with the methods and hence their suitability for different practical applications.

Physical-technical image quality parameters such as noise, resolution and contrast are objective methods for assessment of IQ based on measurements of image data sets. They are essential for purposes as equipment design, performance specification and acceptance and constancy testing [65]. These measurements are not affected by human perception. Consequently, they do not suffer from variations associated with human observers and are

therefore highly reproducible. This is beneficial for studies which are aimed at detecting drifts in equipment performance such as routine quality control [28].

Correspondingly, the performance of CT scanners is frequently measured using physical phantoms. Their performance evaluations are used to perform quality control tests, develop clinical protocols, accredit devices, evaluate clinics for inclusion in clinical trials, or assess the utility of new scanner designs and algorithms. Currently, a number of useful phantoms exist that are targeted to the measurement of image noise, CNR, spatial resolution, Hounsfield Unit accuracy and alignment. Even though these measurements are of significant value for FBP algorithms, they fail to capture the performance of IR algorithms, which are common dose reducing algorithms on modern CT systems. Because these iterative algorithms are nonlinear, the reconstructed resolution and noise can depend on the object contrast, making the standard contrast and noise metrics of limited utility. Additionally, iterative reconstruction algorithms can influence image texture, which these conventional metrics do not capture [66]. Therefore, new parameters such as NPS_{task} are developed to enable the comparison of CT performance in terms of resolution and noise [12].

A very important advantage of physical-technical image quality performance testing is that they are fairly easy to implement in routine practice. These measurements can easily be performed on a regular basis by an in-house medical physics expert, ensuring a thorough evaluation of the obtained image quality of all radiological equipment in a medical imaging department.

The above mentioned physical-technical image quality parameters, while robust and quantitative for physical-technical image quality assessment, are not applicable to clinical images. A system with excellent noise properties but with a poor resolution will not result in an increase of diagnostic performance. Similarly, depending on the diagnostic task, images with good resolution but without decent low contrast detectability will not necessarily result in a good clinical image quality score.

In chapter 7, noise related parameters (artifact area and artifact index) are used to assess the efficiency of different metal artifact reduction tools. These parameters quantify the amount of artifacts present in an image. However, this study shows that the resolution plays a critical role to quantify the effect of metal artifact reduction tools. If the efficiency of such reduction tools would be solely assessed based on noise related parameters, deterioration of the resolution would introduce secondary artifacts compromising diagnostic suitability of the reconstructed image.

In clinical images, complex and variable anatomical structures complicate any quantitative assessment of image quality. Established techniques such as receiver operating characteristic (ROC) or model observers are suitable methods for the assessment of clinical image quality in terms of diagnostic performance [67, 68].

Before a realistic assessment of the clinical image quality can be made, the requirements need to be defined. It is that the required clinical information is contained in the images and can be interpreted by the observer that is important, rather than whether the appearance of the image is pleasing to the eye [69]. The ideal set of parameters describing the clinical image quality should give a measure of the effectiveness with which an image can be used for its intended purpose, namely answering the clinical questions posed. They should therefore relate to the ability of the image to demonstrate disease and to delineate anatomical structures which are relevant to detection, differential diagnosis and localization. The image quality requirements vary for different types of examinations or even different tasks within single examinations [69]. Various imaging tasks require different levels of image quality; an image may be of sufficient quality for one task, but inadequate for another task [70].

One of the challenges radiologists face when assessing new image reconstruction techniques as they are applied to lower dose images, is how to evaluate whether quantitative improvements in noise lead to equivalent or even improved image quality. Qualitative assessment often depends on reader preference, comfort level with image appearance, personality (i.e. the ability to deal with changes in image appearance), and years of experience, rather than a true assessment of quality [71].

The best methodology for image quality optimization studies is definitely based on clinical image quality assessment by means of task specific assessment. However, these studies are difficult to implement in routine practice since such studies are expensive and logistically challenging [72]. Furthermore, the inherent fluctuations in a human observer imply that all methods that involve human observer have limited reliability, meaning that a large number of observations are usually needed in order to obtain reliable result [73]. The use of observer models could overcome this problem. However, as discussed in paragraph 8.3, the implementation of observer models in a medical imaging department requires a profound knowledge of the theoretical background of the different signals, backgrounds and observer models.

Final conclusions

The objectives of this PhD thesis were to assess the reliability and validity of different image quality assessment methods and to investigate a correlation between physical-technical and clinical image quality. This assessment was performed for planar chest radiography, chest CT, and for the efficiency of different metal artifact reduction tools.

In chest radiography, a significant correlation $r = 0.92$ ($p < 0.001$) was found between physical-technical and clinical image quality by means of respectively a contrast detail study and a visual grading analysis. The results indicate that the assessment of the normal anatomy in a chest radiograph is highly correlated with the performance of contrast detail studies. However, this does not automatically imply that a contrast detail analysis is a suitable method to predict the results from ROC studies, since in such studies, pathological structures instead of the normal anatomy have to be assessed. Nevertheless, the excellent correlation between the normal anatomy and contrast detail analysis suggests the value of the latter technique, which is important as contrast detail analysis is a frequently used image quality optimization method.

Although a very good correlation between physical-technical and clinical image quality was found for planar chest radiography, this correlation was less obvious for chest CT images. The image quality for chest CT investigations was assessed using physical-technical parameters like noise, CNR and IQF_{inv} , while clinical image quality was assessed using a visual grading analysis. Based on the different image quality parameters, potential dose reductions were determined for different strengths of iterative reconstruction. Comparison of the obtained potential dose

reductions showed that IQF_{inv} , derived from the contrast-detail curve, correlated best with the clinical image quality. Noise and noise related parameters, i.e. CNR, showed to be not suitable for assessing IR algorithms.

Likewise, the inability of noise and noise related parameters to evaluate image quality was supported in the third study in which the efficiency of different metal artifact reduction tools was investigated. Deterioration of the resolution, which is not captured by noise related image quality parameters, could introduce secondary artifacts compromising diagnostic suitability of the reconstructed image.

Both physical-technical and clinical image quality parameters are very important tools to assess the image quality. However, it is important to understand the strengths and weaknesses of both methods. Physical technical measurements are fairly easy to conduct in quality assurance programs. They are essential for evaluating equipment performance and acceptance and constancy testing. However, for information about the actual clinical performance related to pathological anatomy like lesion detection, clinical assessment studies based on ROC studies still need to be performed. Clinical image quality assessment is rather difficult to perform in routine practice. However, they are the only reliable option for optimization purposes of radiologic diagnosis of specific pathologies.

Based on the results obtained in present thesis with respect to the correlation between physical-technical and clinical image quality, a contrast-detail study is the physical-technical tool closest related to clinical image quality for the assessment of the normal anatomy.

This thesis has certainly proven that image quality analysis is a very complex but very important item. However, in routine quality control and quality assurance programs, the main focus is still on technical exposure parameters like dose and CTDI. There is significantly less attention for image quality parameters. The physical-technical image quality parameters discussed in this thesis can definitely contribute to a thorough analysis necessary for routine image quality assurance. The role of a medical physics expert in this context is crucial. Not only for the correct interpretation of the obtained data but also for ensuring corrective actions if necessary.

The use of Thiel embalmed cadavers proved to be suitable for the assessment of clinical image quality. For both chest RX and chest CT acquisitions and for head CT acquisitions, equivalency

of real patient images and Thiel images was confirmed, indicating the value of Thiel embalmed cadavers in medical X-ray imaging.

Future perspectives

In this PhD thesis, an excellent correlation was found between contrast-detail and VGAS in chest radiography. However, this correlation is not yet investigated for mammography. The CDMAM phantom is an intensively used phantom for image quality analysis during constancy and acceptance testing in mammography. Based on the results of CDMAM analysis, clinical exposure parameters are optimized. This optimization process is based on the assumption that the results of a contrast-detail analysis are reflected on the clinical image quality. Consequently, it is important to establish the correlation between CDMAM analysis and VGAS.

Physical-technical image quality assessment in CT is challenged by the introduction of new reconstruction algorithms. The non-linearity of these algorithms makes it impossible to assess physical-technical image quality with the conventional parameters like noise, NPS and MTF. Therefore, new phantoms which allow more task specific physical-technical image quality assessment need to be developed.

The phantoms used for contrast-detail analysis in digital radiography and mammography contain contrast and detail levels varying over large scales. However, in CT, the current commercially available contrast-detail phantoms contain only a limited amount of contrast levels. This confines the value of these phantoms for image quality assessment in CT images. Since excellent correlation was found between contrast-detail analysis and the normal human anatomy for digital chest radiography, it would be interesting to further explore this correlation in CT images. For this, construction of more sophisticated contrast-detail phantoms is essential.

Thiel embalmed cadavers proved to be a great tool for the assessment of clinical image quality. In all three studies, equivalency of real patient images and Thiel images was confirmed. Thiel embalmed cadavers have the advantage that they can be exposed to multiple acquisitions without ethical problems. Therefore, Thiel embalmed cadavers can be used in future clinical optimization studies of all medical imaging modalities using X-rays.

References

1. Samei, E., W.R. Eyler, and L. Baron, *Handbook of Medical Imaging. Vol 1, Physics and Psychophysics: Effects of anatomical structure on signal detection*. 2000, Bellingham, Washington: SPIE Press.
2. Eckstein, M.P. and J.S. Whiting, *Why do anatomic backgrounds reduce lesion detectability?* Investigative Radiology, 1998. **33**(4): p. 203-208.
3. Samei, E., M.J. Flynn, and W.R. Eyler, *Detection of subtle lung nodules: Relative influence of quantum and anatomic noise on chest radiographs*. Radiology, 1999. **213**(3): p. 727-734.
4. Burgess, A.E., F.L. Jacobson, and P.F. Judy, *Human observer detection experiments with mammograms and power-law noise*. Medical Physics, 2001. **28**(4): p. 419-437.
5. Bochud, F.O., et al., *Estimation of the noisy component of anatomical backgrounds*. Medical Physics, 1999. **26**(7): p. 1365-1370.
6. Vaishnav, J.Y., et al., *Objective assessment of image quality and dose reduction in CT iterative reconstruction*. Medical Physics, 2014. **41**(7).
7. Richard, S., et al., *Towards task-based assessment of CT performance: System and object MTF across different reconstruction algorithms*. Medical Physics, 2012. **39**(7): p. 4115-4122.
8. Chen, B.Y., S. Richard, and E. Samei, *Relevance of MTF and NPS in Quantitative CT: Towards Developing a Predictable Model of Quantitative Performance*. Medical Imaging 2012: Physics of Medical Imaging, 2012. **8313**.
9. Chen, B.Y., et al., *Evaluating iterative reconstruction performance in computed tomography*. Medical Physics, 2014. **41**(12): p. 11.
10. Willemink, M.J., et al., *Iterative reconstruction techniques for computed tomography part 2: initial results in dose reduction and image quality*. European Radiology, 2013. **23**(6): p. 1632-1642.
11. Solomon, J. and E. Samei, *Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE*. Medical Physics, 2014. **41**(9): p. 12.
12. Chen, B.Y., et al., *Assessment of volumetric noise and resolution performance for linear and nonlinear CT reconstruction methods*. Medical Physics, 2014. **41**(7): p. 12.
13. Brady, S.L., et al., *Pediatric CT: Implementation of ASIR for Substantial Radiation Dose Reduction While Maintaining Pre-ASIR Image Noise*. Radiology, 2014. **270**(1): p. 223-231.
14. Winklehner, A., et al., *Raw data-based iterative reconstruction in body CTA: evaluation of radiation dose saving potential*. European Radiology, 2011. **21**(12): p. 2521-2526.
15. Pickhardt, P.J., et al., *Abdominal CT With Model-Based Iterative Reconstruction (MBIR): Initial Results of a Prospective Trial Comparing Ultralow-Dose With Standard-Dose Imaging*. American Journal of Roentgenology, 2012. **199**(6): p. 1266-1274.
16. Willemink, M.J., et al., *Computed Tomography Radiation Dose Reduction: Effect of Different Iterative Reconstruction Algorithms on Image Quality*. Journal of Computer Assisted Tomography, 2014. **38**(6): p. 815-823.
17. Dobbins, J.T., *Handbook of Medical Imaging. Vol 1, Physics and Psychophysics: Image quality metric for digital system*. 2000, Bellingham, Washington: SPIE Press.
18. Schindera, S.T., et al., *Iterative Reconstruction Algorithm for CT: Can Radiation Dose Be Decreased While Low-Contrast Detectability Is Preserved?* Radiology, 2013. **269**(2): p. 510-517.

19. Rose, A., *The sensitivity performance of the human eye on an absolute scale*. J Opt Soc Am, 1948. **38**(2): p. 196-208.
20. Honey, I.D., A. MacKenzie, and D.S. Evans, *Investigation of optimum energies for chest imaging using film-screen and computed radiography*. British Journal of Radiology, 2005. **78**(929): p. 422-427.
21. Hamer, O.W., et al., *Contrast-detail phantom study for X-ray spectrum optimization regarding chest radiography using a cesium iodide - Amorphous silicon flat-panel detector*. Investigative Radiology, 2004. **39**(10): p. 610-618.
22. Bacher, K., et al., *Image Quality and Radiation Dose on Digital Chest Imaging: Comparison of Amorphous Silicon and Amorphous Selenium Flat-Panel Systems*. American Journal of Roentgenology, 2006. **187**: p. 630-637.
23. Commission, E., *The European protocol for the quality control of the physical and technical aspects of mammography screening: part B. Digital mammography*, in *European Guidelines for Breast Cancer Screening*. 2006: Luxembourg.
24. Van Metter, R., M. Heath, and L. Fletcher-Heath, *Applying the "European protocol for the quality control of the physical and technical aspects of mammography screening" threshold contrast visibility assessment to digital systems*. Medical Imaging 2006: Physics of Medical Imaging, Pts 1-3, 2006. **6142**.
25. Cohen, G., D.L. McDaniel, and L.K. Wagner, *Analysis of Variations in Contrast-Detail Experiments*. Medical Physics, 1984. **11**(4): p. 469-473.
26. Artinis Medical Systems, *Manual CDRAD 2.0 Phantom & analyser software version 2.1*.
27. Thijssen, M. and K.R. Bijkerk, *Manual Contrast-detail phantom Artinis CDMAM type 3.4*, Artinis Medical Systems. Department of Radiology, University Medical Center Nijmegen, The Netherlands.
28. Pascoal, A., et al., *Evaluation of a software package for automated quality assessment of contrast detail images - comparison with subjective visual assessment*. Physics in Medicine and Biology, 2005. **50**(23): p. 5743-5757.
29. Tingberg, A., et al., *Evaluation of image quality of lumbar spine images: A comparison between FFE and VGA*. Radiation Protection Dosimetry, 2005. **114**(1-3): p. 53-61.
30. Zanca, F., et al., *Comparison of visual grading and free-response ROC analyses for assessment of image-processing algorithms in digital mammography*. British Journal of Radiology, 2012. **85**(1020): p. E1233-E1241.
31. Commission of the European Communities, *European guidelines on quality criteria for diagnostic radiographic images (EUR 16260)*. 1996.
32. Tapiovaara, M.J., *STUK-A219: Relationships between physical measurements and user evaluation of image quality in medical radiology - a review*. 2006, STUK.
33. Bath, M. and L.G. Mansson, *Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation*. British Journal of Radiology, 2007. **80**(951): p. 169-176.
34. Tingberg, A., et al., *Comparison of two methods for evaluation of the image quality of lumbar spine radiographs*. Radiation Protection Dosimetry, 2000. **90**(1-2): p. 165-168.
35. Burgess, A.E., et al., *EFFICIENCY OF HUMAN VISUAL SIGNAL DISCRIMINATION*. Science, 1981. **214**(4516): p. 93-94.
36. Abbey, C.K. and M.P. Eckstein, *Estimates of human-observer templates for a simple detection task in correlated noise*, in *Medical Imaging 2000: Image Perception and Performance*, E.A. Krupinski, Editor. 2000, Spie-Int Soc Optical Engineering: Bellingham. p. 70-77.

37. Bochud, F.O., C.K. Abbey, and M.P. Eckstein, *Statistical texture synthesis of mammographic images with clustered lumpy backgrounds*. Optics Express, 1999. **4**(1): p. 33-43.
38. Reiser, I. and R.M. Nishikawa, *Identification of simulated microcalcifications in white noise and mammographic backgrounds*. Medical Physics, 2006. **33**(8): p. 2905-2911.
39. Eckstein, M.P., C.K. Abbey, and J.S. Whiting, *Human vs. model observers in anatomic backgrounds*. Image Perception, 1998. **3340**: p. 16-26.
40. Zhang, Y., B. Pham, and M.P. Eckstein, *Evaluation of JPEG 2000 encoder options: Human and model observer detection of variable signals in X-ray coronary angiograms*. Ieee Transactions on Medical Imaging, 2004. **23**(5): p. 613-632.
41. Veldkamp, W.J.H., et al., *A Technique for Simulating the Effect of Dose Reduction on Image Quality in Digital Chest Radiography*. Journal of Digital Imaging, 2009. **22**(2): p. 114-125.
42. Castella, C., et al., *Human linear template with mammographic backgrounds estimated with a genetic algorithm*. Journal of the Optical Society of America a-Optics Image Science and Vision, 2007. **24**(12): p. B1-B12.
43. Bochud, F.O., C.K. Abbey, and M.P. Eckstein, *Visual signal detection in structured backgrounds. III. Calculation of figures of merit for model observers in statistically nonstationary backgrounds*. Journal of the Optical Society of America a-Optics Image Science and Vision, 2000. **17**(2): p. 193-205.
44. Zhang, Y., B.T. Pham, and M.P. Eckstein, *Evaluation of internal noise methods for Hotelling observer models*. Medical Physics, 2007. **34**(8): p. 3312-3322.
45. Barrett, H.H., et al., *Model Observers for Assessment of Image Quality*. Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(21): p. 9758-9765.
46. Leng, S., et al., *Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain*. Medical Physics, 2013. **40**(8).
47. Eckstein, M.P., C. Abbey, and F. Bochud, *Handbook of Medical Imaging. Vol 1, Physics and Psychophysics: A practical guide to model observers for visual detection in synthetic and natural noisy images*. 2000, Bellingham, Washington: SPIE Press.
48. Barrett, H.H., *Objective Assessment of Image Quality - Effects of Quantum Noise and Object Variability*. Journal of the Optical Society of America a-Optics Image Science and Vision, 1990. **7**(7): p. 1266-1278.
49. Strotzer, M., et al., *Detection of simulated chest lesions with normal and reduced radiation dose - Comparison of conventional screen-film radiography and a flat-panel x-ray detector based on amorphous silicon*. Investigative Radiology, 1998. **33**(2): p. 98-103.
50. Goo, J.M., et al., *Detection of simulated chest lesions by using soft-copy reading: Comparison of an amorphous silicon flat-panel-detector system and a storage-phosphor system*. Radiology, 2002. **224**(1): p. 242-246.
51. Albertini, F., et al., *Experimental verification of IMPT treatment plans in an anthropomorphic phantom in the presence of delivery uncertainties*. Physics in Medicine and Biology, 2011. **56**(14): p. 4415-4431.
52. Samei, E., et al., *Subtle Lung Nodules: Influence of Local Anatomic Variations on Detection*. Radiology, 2003. **228**: p. 76-84.
53. Pattanshetti, V.M. and S.V. Pattanshetti, *Laparoscopic surgery on cadavers: a novel teaching tool for surgical residents*. Anz Journal of Surgery, 2010. **80**(10): p. 676-678.
54. Holzle, F., et al., *Thiel embalming technique: a valuable method for teaching oral surgery and implantology*. Clin Implant Dent Relat Res, 2012. **14**(1): p. 121-6.

55. Wolff, K.D., et al., *Thiel embalming technique: a valuable method for microvascular exercise and teaching of flap raising*. Microsurgery, 2008. **28**(4): p. 273-8.
56. Benkhadra, M., et al., *Comparison of fresh and Thiel's embalmed cadavers according to the suitability for ultrasound-guided regional anesthesia of the cervical region*. Surgical and Radiologic Anatomy, 2009. **31**(7): p. 531-535.
57. Healy, S.E., et al., *Thiel Embalming Method for Cadaver Preservation: A Review of New Training Model for Urologic Skills Training*. Urology, 2015.
58. Chevallier, C., et al., *Postmortem Circulation : A New Model for Testing Endovascular Devices and Training Clinicians in Their Use*. Clinical Anatomy, 2014. **27**(4): p. 556-562.
59. Hammer, N., et al., *Comparison of Modified Thiel Embalming and Ethanol-Glycerin Fixation in an Anatomy Environment: Potentials and Limitations of Two Complementary Techniques*. Anatomical Sciences Education, 2015. **8**(1): p. 74-85.
60. Rai, B.P., et al., *A qualitative assessment of human cadavers embalmed by Thiel's method used in laparoscopic training for renal resection*. Anatomical Sciences Education, 2012. **5**(3): p. 182-186.
61. International Commission on Radiation Units and Measurements, *ICRU Report 41: Modulation transfer function of screen-film systems*. 1996.
62. Sandborg, M., et al., *Demonstration of correlations between clinical and physical image quality measures in chest and lumbar spine screen-film radiography*. British Journal of Radiology, 2001. **74**(882): p. 520-528.
63. Sandborg, M., et al., *Comparison of clinical and physical measures of image quality in chest and pelvis computed radiography at different tube voltages*. Medical Physics, 2006. **33**(11): p. 4169-4175.
64. Redlich, U., C. Hoeschen, and W. Doehring, *Assessment and optimisation of the image quality of chest-radiography systems*. Radiation Protection Dosimetry, 2005. **114**(1-3): p. 264-268.
65. Tapiovaara, M., *Image quality measurements in radiology*. Radiation Protection Dosimetry, 2005. **117**(1-3): p. 116-119.
66. Wilson, J.M., et al., *A methodology for image quality evaluation of advanced CT systems*. Medical Physics, 2013. **40**(3).
67. Busch, H.P. and K. Faulkner, *Image quality and dose management in digital radiography: A new paradigm for optimisation*. Radiation Protection Dosimetry, 2005. **117**(1-3): p. 143-147.
68. Lin, Y., et al., *An image-based technique to assess the perceptual quality of clinical chest radiographs*. Medical Physics, 2012. **39**(11): p. 7019-7031.
69. Martin, C.J., P.F. Sharp, and D.G. Sutton, *Measurement of image quality in diagnostic radiology*. Applied Radiation and Isotopes, 1999. **50**(1): p. 21-38.
70. Maidment, A.D.A., *Diagnostic Radiology Physics: A handbook for teachers and students: Measures of image quality*. 2014, Vienna, Austria: International Atomic Energy Agency.
71. Baker, M.E., et al., *Contrast-to-Noise Ratio and Low-Contrast Object Resolution on Full- and Low-Dose MDCT: SAFIRE Versus Filtered Back Projection in a Low-Contrast Object Phantom and in the Liver*. American Journal of Roentgenology, 2012. **199**(1): p. 8-18.
72. Samei, E. and S. Richard, *Assessment of the dose reduction potential of a model-based iterative reconstruction algorithm using a task-based performance metrology*. Medical Physics, 2015. **42**(1): p. 314-323.
73. Bath, M., *Evaluating imaging systems: practical applications*. Radiat Prot Dosimetry, 2010. **139**(1-3): p. 26-36.

Part V:

Curriculum Vitae: An De Crop

Curriculum Vitae

An De Crop

PERSONAL INFORMATION

Address: Kraenepoelpad 11, 9880 Aalter
E-mail: andecrop@gmail.com
Mobile: +32 (0)472 36 83 65
Date of birth: July 31, 1984
Place of birth: Eeklo, Belgium
Nationality: Belgian
Civil status: Married

PROFESSIONAL EXPERIENCE

- 2008 – present:
Teaching and research assistant, Ghent University
Medical Physicist, QCC-Gent, Ghent University
PhD student, Dept. of Basic Medical Sciences, Ghent University
- 2010 – present:
Lecturer Radiation protection, University College Ghent
- Licensed expert in medical physics (radiology) since Sep 2008.

EDUCATION

- 2008 – 2015:
PhD student in medical sciences, Ghent University
Thesis: Image quality evaluation in X-ray medical imaging based on Thiel embalmed human cadavers
Supervisors: Prof. Dr. H. Thierens and Prof. Dr. K. D'Herde

- 2006 – 2008:
Postgraduate studies expert in medical radiation physics (radiology), Ghent University
Thesis: Effective dose contribution of cone-beam CT acquisition during uterine artery embolization
Supervisors: Prof. Dr. H. Thierens and Prof. Dr. ir. K. Bacher
- 2002 – 2006:
Master in biomedical sciences, Ghent University
Thesis: Patient radiation dose for radiological follow-up of laparoscopic adjustable gastric banding
Supervisors: Prof. Dr. H. Thierens and Prof Dr. Ir K. Bacher

A1 PUBLICATIONS

- **De Crop A**, Casselman J, Van Hoof T, et al. Analysis of metal artifact reduction tools for dental hardware in CT scans of the oral cavity: kVp, Iterative Reconstruction, Dual Energy CT, Metal Artifact Reduction Software: Does it make a difference? *Neuroradiology* (2015): epub ahead of print.
- **De Crop A**, Smeets P, Van Hoof T, et al. Correlation of clinical and physical-technical image quality in chest CT: a human cadaver study applied on iterative reconstruction. *BMC Medical Imaging* (2015): 15(1), 32.
- **De Crop A**, Bacher K, Van Hoof T, et al. Correlation of Contrast-Detail Analysis and Clinical Image Quality Assessment in Chest Radiography with a Human Cadaver Study. *Radiology* (2012):26 (1),298-304

CONFERENCES

INTERNATIONAL CONFERENCES

- Annual meeting of the Radiological Society of North America: November 2012, Chicago, USA
- European Congress of Radiology : March 2011, Vienna, Austria
Oral presentation: Correlation of contrast-detail analysis and clinical image quality for chest radiography based on a human cadaver study. **De Crop A**, Bacher K, Van Hoof T, Smeets P, D'Herde K, Verstraete K, Thierens H.
Poster presentation: Evaluation of a new instrument for on-site dose area product meter calibration. K. Bacher, **A. De Crop**, H. Thierens

- World Congress on Medical Physics and Biomedical Engineering: September 2009, Munich, Germany.
Oral presentation: Effective Dose Contribution of Cone-Beam CT Acquisition during Uterine Artery Embolization: A Phantom Study. **A. De Crop**, K. Bacher, L. Defreyne, and H. Thierens. IFMBE Proceedings 25/III, p. 455 - 457.

NATIONAL CONFERENCES

- XXX Annual Symposium of the Belgian Hospital Physicists Association: February 2015, Antwerp, Belgium
- XXVI Annual Symposium of the Belgian Hospital Physicists Association: February 2011, Chaleroi, Belgium
Oral presentation: Correlation of contrast-detail analysis and clinical image quality for chest radiography based on a human cadaver study. **De Crop A**, Bacher K, Van Hoof T, Smeets P, Duyck Ph, Verstraete K, D'Herde K, Thierens H.
- XXV Annual Symposium of the Belgian Hospital Physicists Association: February 2010, Brussels, Belgium
Oral presentation: Dose area product measurements in dental panoramic radiography: an accuracy assessment of four different methods. **A. De Crop**, E. Pauwels, L. Vereecken, A. De Hauwere, H. Thierens, K. Bacher.
Poster presentation: K. Bacher, **A. De Crop**, H. Thierens. Evaluation of a new instrument for on-site dose area product meter calibration.
- XXIV Annual Symposium of the Belgian Hospital Physicist Association: February 2009, Courtrai, Belgium
Oral presentation: Effective dose contribution of cone-beam CT acquisition during uterine artery embolization: a phantom study. **A. De Crop**, K. Bacher, L. Defreyne, H. Young physicist award

